1997

Applied academics: an evaluation of the applied academics program in Iowa

Dennis Wayne Field
Iowa State University

Applied academics:

An evaluation of the applied academics program in Iowa

by

Dennis Wayne Field

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Industrial Education and Technology

Major Professor: John C. Dugger, III

Iowa State University

Ames, Iowa

1997

UMI Number: 9725412

Graduate College

Iowa State University

This is to certify that the Doctoral dissertation of

Dennis Wayne Field

has met the dissertation requirements of Iowa State University

Signature was redacted for privacy.

Major Professor

Signature was redacted for privacy.

For the Major Program

Signature was redacted for privacy.

For the Graduate College

DEDICATION

To Susan:

Your love, patience, and support have meant more to me than you know. I could not have done this without you.

To Katie and Matt:

Your mother and I were truly blessed when you entered our lives.

To the memory of Doris Elaine Field:

With our family in spirit always.

TABLE OF CONTENTS

ABSTRACT

This investigation compares academic achievement and employability skills of high school students enrolled in applied academics courses versus traditional courses. Outcomes from three American College Testing Work Keys assessment tests--Applied Mathematics (AM), Applied Technology (AT), and Reading for Information (RFI)--were used as measures of employability skills. Data were collected under quasi-experimental conditions on 1,321 students from 9 Iowa high schools. The data included school, type of course (applied or traditional), course subject matter, class within course, gender, grade, grade point average (GPA), Iowa Tests of Educational Development (ITED) score, test content area, and test score.

Findings included:

- Group means for GPA, ITED, and all 3 tests were higher for traditional than applied students.
- Students scoring below the minimum skill level assessed on the tests (Level 3) were not restricted to those with below average GPA or ITED scores.
- Over 41% of students taking the AT test scored below the cutoff of 3. In contrast only 7.2% of all students taking the RFI test and only 2.5% of all students taking the AM test scored below the cutoff.

- Males had higher average scores on the AM and AT tests, while females had higher average scores on the RFI test.

Conclusions and recommendations include:

- These findings should not be taken as evidence of a superiority of traditional teaching methods over applied academics. These were not equivalent groups being compared under true experimental conditions; nor can one discount the possibility of omitted intercorrelated independent variables in the regression equations.
- Additional measures of employability skills besides test scores are needed to fully investigate the effectiveness of applied academics.
- Growth of students' employability skills over time should be monitored. Data should be collected at periodic intervals for analysis and should include measures of performance in both school and workplace.
- Independent variables, other than those included in the study, may account for significant variability related to test scores.
- The results of this investigation should not be generalized to all Iowa high schools due to limited sample size and lack of an adequate cross-section.

CHAPTER 1. INTRODUCTION

Rationale

A uniquely American philosophy of education arose from the needs of the new nation and has been described by one author (Kandel, 1958) as follows:

> It was recognized early in the history of the Republic that education must be devoted to the unification of the American people on the basis of democratic ideals. Equality of educational opportunities was to be provided for all. Instruction at all levels of the educational system was to be directed to suit the capacities of each pupil and to contribute to the economic progress of the nation by providing the training necessary to make the best uses of the country's vast resources. Above all, however, the end of education was to train for intelligent citizenship and participation in the affairs of government. These aims and particularly the last one mentioned have been stated and reaffirmed by all leaders of American life from the days of George Washington to the present. (pp. 18-19)

As one might expect when multiple objectives are set forth, as they are above, and resources to accomplish these objectives are limited, conflicts inevitably arise. When one examines the evolution of the present American educational system and the forces that played a part in it, one may notice the appearance of competing philosophies and purposes at various junctures. Resnick and Wirt (1996) summarize the struggle between the Jeffersonian ideal of a universally educated yeoman citizenry and the needs of an emerging industrial society:

> From the earliest years of public education in America, leading educators--Horace Mann in the nineteenth century and John Dewey in the twentieth, for example--aimed for schools that would cultivate the questioning and reasoning processes and the skills of democratic social interaction that were needed by all citizens in a properly functioning democracy. Others joined with the democratic theorists to promote education for full personal

> lives, to encourage the lifelong learning and the capacity to engage with enthusiasm and competence in the multiple pursuits, from parenting to leisure activities, that would fill people's longer and longer lives. But the demands of the growing industrial economy were different. Industrialists called for a large supply of literate but essentially docile factory workers who would accept the boring and sometimes dangerous conditions of industrial production. Their view of education was locked into place early in this century by a series of policy and educational management decisions that modeled American school systems on the efficient, Taylorized factory. (p. 9)

Although a less-than-flattering portrait of industrialists is painted above, one could argue effectively that in this country there is, and always has been, an expectation that educational systems address the need to prepare individuals for work. The use of the educational system to meet the economic need of a society is not an American invention. Bennett (1926) notes, for example, that this has been a "fundamental motive" since antiquity. In his classic on the history of industrial education, he states that:

> ... the ancient Jews recognized that to fail to give a boy an honest means of livelihood, which usually meant giving him instruction in some manual trade, was to prepare him to be a social parasite, dangerous to the community. On the other hand, to make him skilful [sic] in a manual trade was to insure his becoming a useful member of society. (p. 13)

While his writings lack some of the cultural sensitivity we would expect today, his findings are no less valid; education of all members of society is in the national interest. As Hartoonian and Van Scotter (1996) point out: It is a fact that when economic hard times are upon us, more often than not, a finger of blame is pointed in the direction of the educational system.

> In the closing decades of the 19th century, authorities criticized schools as the United States continued to lose market share to the Germans in the

> machine tool industry. Then in the 1930s, educators were told that, if schools had better educated students for employment, we might have avoided the Great Depression. ... And in the wake of the economic recession of the late 1950s, assessment of schooling was linked to the performance of the economy. (p. 556)

As we found ourselves buffeted by competition from all over the world in the '70s, '80s, and '90s, we again looked toward our educational systems. In an industrialized, high-wage economy such as the U.S., global competitiveness depends to a great extent on skill level of the pool of workers. A report of the Commission on the Skills of the American Workforce entitled America's Choice: High Skills or Low Wages! (National Center on Education and the Economy, 1990) noted that the U.S. must invest more in educating and training our workforce if we hope to effectively compete with high-skill economies. According to Boesel and McFarland (1994), the time when the US could maintain a competitive edge in low skill, labor intensive businesses is clearly past:

> Our competitors also include nations with less well educated but disciplined workforces able to perform the sort of semi-skilled work that has been the backbone of American manufacturing, and willing to do so for lower wages. Thus, many American manufacturing jobs have migrated to countries such as Taiwan, Korea, and Mexico. (p. 9)

It is in the best interest of the United States to develop and maintain a highly skilled pool of workers. The difficulty lies in how best to achieve that goal. The answer is not to produce ever-higher numbers of college graduates as Carson, Huelskamp and Woodall (1993) report. They reference the findings of researchers at the Sandia National Laboratories who maintain that: "The education system turns out in today's youth roughly 26% as college graduates, an additional 60% with 12 to 15 years of schooling, and the final 14% with less than a high school diploma" (pp. 293-294). These researchers claim that the above percentages match up fairly well with what employers need. The same researchers state that the numbers are also consistent with the results from two other studies: the Hudson Institute's Workforce 2000 report and the above-mentioned report entitled America's Choice: High Skills or Low Wages!. The estimate regarding the need for employees with college degrees is further substantiated by Gray (1996) who notes that:

> The reality is that, since the 1950s, only around 30% of all jobs have required a four-year college degree and only 20% of all employment has been in the professional ranks. These ratios are not predicted to change in the future. (p. 530)

Given the above percentages and the goal of developing a "high-skill" workforce, it would seem that one could make the most inroads by concentrating educational reform efforts on that 60% of the population whose formal education will end at, or within a few years of high school graduation. There is an emphasis on reform because of the real concern on the part of many groups that educational models now in use at the high school level are inadequate to meet current and future societal needs. Gray (1996) makes the case that our current system of education focuses primarily on the academically gifted student; students who have the greatest chance of success at the university level.

> In most college-prep classes, students are expected to act like office copying machines: the teacher lectures, and the students take notes and then reproduce on the test what they copied. While learning experts argue that this is the least effective teaching strategy for all students, it is

> mastered early on by the academically blessed, who excel as the content becomes more abstract and more detached from any context. The problem is that the academically blessed now amount to less than one-third of those in the college-prep curriculum. The majority come from the academic middle and do not learn this way very well.
>
> The learning styles of those in the academic middle are typically more concrete. They learn best when instruction is put into relevant "real world" context. (pp. 533-534)

Gray goes on to say, speaking of those students between the 25th and 75th percentiles of the academic continuum, that:

> As things now stand, students from the academic middle typically show low levels of academic engagement. ... Being neither gifted nor handicapped, they do not fit into legally defined categories and, therefore, receive little attention and have few advocates. Despite the fact that they are now found primarily in the college-prep programs, they remain invisible ... . [This] is both unfair and counterproductive to national interests. (p. 534)

If what Gray says is true, then we could expect widespread dissatisfaction with the results of the educational process for this group in the academic middle. There is substantial evidence that just such widespread dissatisfaction exists, although it extends to more than just the academic performance of the students. Personal qualities such as responsibility, integrity, self-management, sociability, self-esteem, and honesty are also needed to meet job performance expectations per the Secretary's Commission on Achieving Necessary Skills (1991, p. vii).

In a report to Congress through the Office of Educational Research and Improvement, Boesel and McFarland (1994) mention that "a significant amount of vocational education, particularly secondary vocational education, has failed to respond to the emerging skill needs of employers" (p. 1).

In Iowa, concerns regarding education and its effect on workforce readiness were expressed in a report released by the ACT Center for Education and Work (1995). In it, the authors state that, "Employers, frustrated, have vented to educators that curricula are not sufficiently focused on preparing students for work. Today's employees, they say, must have higher skill levels in communications, mathematics, and teamwork" (p. 5). This report, titled Making the Grade: Keys to Success on the Job in the 90's was prepared at the request of The Iowa Business Council (IBC).

Finally, the SCANS report contains the following statement:

> ... more than half of our young people leave school without the knowledge or foundation required to find and hold a good job. Unless all of us work to turn this situation around, these young people, and those who employ them, will pay a very high price. Low skills lead to low wages and low profits. Many of these youth will never be able to earn a decent living. And, in the long run, this will damage severely the quality of life everyone hopes to enjoy. None of us, and none of you, wants to stand by while this happens. (Secretary's Commission on Achieving Necessary Skills, 1991, p. v)

There is a ground swell for change and the pressure for change is coming from stakeholders in the educational process apart from the classroom. Goldberger and Kazis (1996) summarize the current state of affairs by noting that:

> Much of the recent impetus for improving the transition from school to work has come from outside the schools. Employers concerned about the quality of entry-level employees have been active proponents, as have policy makers and analysts from fields outside education, particularly economic development and employment and training. (p. 547)

This push for change takes many forms at both the federal and state levels; but whatever form it may take, there is consensus for change. An independent advisory panel associated with the July 1994 National Assessment of Vocational Education (NAVE) report stated that:

> Vocational education should be high quality: It should be competency-based, with industry involvement. Industry-oriented skill standards should be used as the mechanism for connecting vocational education to the larger system of education and training. In combination with academic and employability skills, skill standards will provide all students with a rigorous preparation for work and life. (Boesel & McFarland, 1994, p. 1)

The SCANS report counsels employers to:

> ... tell educators clearly what you need and work with them to accomplish it. You know that students have to believe that you care about what they learn. Employers who value performance in high school when they make their hiring decisions provide students with the right signal: learning and earning are related activities.
>
> Finally, ... confirm that the SCANS skills accurately reflect your local workforce requirements. Having confirmed these skills, make sure your local school board never loses sight of them in instructional planning. (Secretary's Commission on Achieving Necessary Skills, 1991, p. viii)

The Iowa Business Council took this recommendation to heart and in 1993 began a project with just such goals in mind. The project is described in the aforementioned report entitled <u>Making the Grade: Keys to Success on the Job in the 90's</u>.

> The council, recognizing the inadequacy with which employers had been communicating to educators, set out to articulate clearly and quantitatively the skills and levels of skills needed by high school graduates to qualify for certain nonexempt positions in its member companies.
>
> The Iowa Business Council engaged American College Testing's (ACT's) Center for Education and Work to help accomplish this goal. Using ACT Work Keys job profiling (job analysis) system, member

> companies of the Iowa Business Council began the process of measuring quantitatively the skills needed for their jobs. The members reasoned that the Work Keys system, which comprises job profiling, work-related assessments, and instructional aids, offered a relatively inexpensive way to measure skills on a statewide basis. ...
>
> The Work Keys System was selected as the catalyst to clarify the link between employment and education. It established a common language for schools and businesses to use to communicate a set of workplace skills and the necessary levels, or standards, for those skills. It also provided a means of measuring these skills both for jobs and for people that could guide learning at the student level ... . (ACT Center for Education and Work, 1995, pp. 5-6)

Herein we have a serious attempt to meet the intent of the SCANS and Office of Educational Research and Improvement recommendations. The first step being the joint development of industry-oriented skill standards by a group of leaders in business and education. The result of this effort is the Work Keys system, which seeks to quantify certain employability skills. The importance of this effort cannot be underestimated. Consider the following quote attributed to Lord Kelvin by Sir William Thomson:

> ... a first essential step in the direction of learning any subject is to find principles of numerical reckoning and methods for practicably measuring some quality connected with it. I often say that when you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre [sic] and unsatisfactory kind: it may be the beginning of knowledge, but you have scarcely, in your thoughts, advanced to the stage of science. (as cited in Merton, Sills, & Stigler, 1984, p. 327)

The second step in an attempt to meet the intent of the SCANS and Office of Educational Research and Improvement recommendations lies in the area of curricula development. There is overwhelming support for the use of applied learning. The SCANS report, for instance, contains the following statement:

> We believe, after examining the findings of cognitive science, that the most effective way of learning skills is "in context," placing learning objectives within a real environment rather than insisting that students first learn in the abstract what they will be expected to apply. (Secretary's Commission on Achieving Necessary Skills, 1991, p. xv)

While the Office of Educational Research and Improvement declares that

> ... there is evidence that "contextualized learning," that is, education in a context that enables students to relate schoolwork to the world outside of school, is a more effective pedagogical approach than traditional education, which emphasizes knowledge for its own sake. There is also evidence from the military that a contextual approach to training results in better job performance than other methods .... (Boesel & McFarland, 1994, p. 35)

Given this overwhelming support for contextual (applied) curricula, it is not surprising that a number of organizations have stepped in to fill the need. These organizations include groups such as the Center for Occupational Research and Development (CORD), the Agency for Instructional Technology (AIT), and the Mid-America Vocational Curriculum Consortium (MAVCC). Other groups, such as the National Tech Prep Demonstration Center at Mt. Hood Community College in Oregon, are also involved with infusion of applied curricula in high schools. (Limback & Rosa, 1996, pp. 150-151)

One might ask at this point how applied curriculum materials are differentiated from traditional materials. The characteristics of applied materials were detailed by Wang and Owens (1994). Their findings indicated that applied materials provide the following features:

- Use modularized student units
- Incorporate teacher-empowering guides
- Include competency-based objectives
- Are enhanced by instructional videos
- Are written at an estimated eighth-grade reading level
- Target secondary vocational students as the primary audience; also useful in postsecondary adult learning settings
- Emphasize holistic learning
- Can be infused into vocational courses or taught alone as a credit course by either vocational or academic instructors--or a team that includes both
- Are not meant to replace "traditional" academic courses for the top 25 percent of the student population
- Emphasize developing teamwork skills in students (as cited in Limback & Rosa, 1996, p. 151)

Iowa schools are no different from many others in the nation. In reaction to the need to strengthen skills of students coming out of high schools, Iowa high schools are implementing curricula changes to address this need. They are supported in this effort by federal legislation. According to some associated with the 1994 NAVE project, all federal education and training legislation should complement and strengthen the system of workforce preparation whose key elements include "standards (academic and industry-linked), assessment, credentialing, curriculum frameworks, teacher training, labor market information, and planned pathways" (Boesel & McFarland, 1994, pp. 2-3).

Various vocational technical programs (School-to-Work, Tech Prep, etc.) have been initiated at both the state and federal levels in an attempt to close the gap between education and employability skills. Applied academics, a component of Tech Prep, is one such effort. Table 1.1 shows the extent to which applied academics programs have been introduced in Iowa high schools.

Table 1.1. Iowa public high school data for the 1995-1996 school year (figures supplied under "Number of High Schools, Applied Academics" refers to schools offering at least one applied academics course during the school year)

	Number of High Schools			Enrollment		
AEA[a]	Total	Applied Academics	% of Total	Total	Applied Academics	% of Total
1	26	26	100%	11,866	2,415	20%
2	23	19	83%	6,977	1,420	20%
3	16	16	100%	4,162	1,435	34%
4	14	12	86%	3,642	988	27%
5	29	24	83%	7,857	2,355	30%
6	15	14	93%	5,160	1,315	25%
7[b]	*22*	*4*	*18%*	*10,191*	*518*	*5%*
8	AEA 8 was previously combined with another AEA-- only 15 "AEAs" exist					
9[b]	*21*	*12*	*57%*	*15,881*	*1,690*	*11%*
10[b]	*33*	*22*	*67%*	*17,456*	*1,644*	*9%*
11[b]	*56*	*24*	*43%*	*32,458*	*2,048*	*6%*
12	22	20	91%	6,979	1,814	26%
13	31	23	74%	10,356	1,347	13%
14	18	18	100%	3,694	959	26%
15	24	14	58%	7,326	765	10%
16	12	10	83%	5,777	1,153	20%
Total	362	258	71%	149,782	21,866	15%

[a] AEA is an acronym for Area Education Agencies. These AEAs function as intermediate units among the Department of Education, school districts, and local schools in Iowa.

[b] *indicates incomplete data*

Unfortunately, most such programs are missing at least some of the key elements needed to measure how effective the programs are in better preparing today's students for the workforce. The recognition of this fact led to the issuance of a Request for Proposals by the Iowa Department of Education to "design and conduct a longitudinal study to determine the effectiveness of the applied academic component of Tech Prep programs being implemented in Iowa schools (H. H. Custer, personal communication, August 31, 1995)." This dissertation is an outgrowth of one segment of that study.

Statement of the Problem

Employers in Iowa and elsewhere perceive a gap between the level of employability skills of students leaving high school and the level needed to obtain and keep a job in most organizations. Many of Iowa's schools are implementing applied academics courses in an attempt to close this gap. While a considerable amount of anecdotal information exists regarding the effectiveness of applied academics courses, the impacts of these curricula changes are not being systematically evaluated to determine if they have the desired effect on student employability skills.

Purpose of the Study

This study was designed with two goals in mind:

1. Compare the academic achievement, on selected variables, of a sample of Iowa high school students enrolled in applied academics courses against those enrolled in traditional courses.

2. Compare the level of selected employability skills--Applied Math, Applied Technology, and Reading for Information--of a sample of Iowa high school students enrolled in applied academics courses against those enrolled in traditional courses.

Research Questions

The reader should note that the research questions are divided into two parts; each related back to one of the two goals previously mentioned.

In order to assess the differences in academic performance between students who have completed applied academics courses and traditional academics courses, the following questions were formulated:

1.1 Is there a statistically significant difference between the high school GPAs of students who have completed applied courses versus comparable traditional courses?

1.2 Is there a statistically significant difference between the composite ITED scores of students who have completed applied courses versus comparable traditional courses?

In order to assess the relationships between applied academics courses and student's employability skills, the following questions were formulated:

2.1 Is there a statistically significant difference in the raw Work Keys assessment test scores for students who have completed applied courses when compared with students who have completed traditional courses?

2.2 Is there a statistically significant correlation with the raw Work Keys assessment test scores and the following concomitant student variables: (a) grade level, (b) grade point average (GPA), (c) composite score on the Iowa Tests of Educational Development (ITED), or (d) gender.

2.3 Is there a statistically significant difference in the adjusted Work Keys assessment test scores for students who have completed applied courses when compared with students who have completed traditional courses? The word "adjusted" indicates that the raw Work Keys assessment test scores have had the effects of significant concomitant variables removed.

Assumptions of the Study

This study was undertaken with the following assumptions:

1. Students put forth their best effort in completing the Work Keys assessment tests and these Work Keys tests are valid and reliable measures of employability skills.
2. The academic outcomes taught in an applied academics course are the same as those covered in the corresponding traditional course and students put forth their best effort during instruction; that is, they try to learn the material presented.
3. Instructors in traditional courses did not introduce any "applied" materials or teaching techniques in their courses, nor did instructors in applied courses revert to traditional instructional methods at any time during the courses.

4. The selection process or statistical "adjustments" resulted in an unbiased sample. In other words, the samples are representative of the target population.
5. Values of the independent variables provided by students, teachers, or administrators were accurate assessments of student characteristics; that is, there were no measurement errors in data provided on such variables as GPA, ITED score, grade level, gender, etc.
6. All necessary statistical assumptions were met for the methods of analysis used in this study. For example, if a particular analysis technique requires that the data be normally distributed, then the data were assumed to be normally distributed.
7. The concomitant variables are not affected by the "treatments" in any way. This would mean, for instance, that any variable that could be affected by a treatment in this study is measured prior to application of the treatment.
8. The data are not biased by variables outside the scope of the study. Such variables would include a student's prior work experience, age of curricula materials, the skill level of the instructor, an instructor's enthusiasm for the course material, socioeconomic status of the students, etc.
9. The study meets conditions necessary for internal and external validity.

Delimitations of the Study

This study is subject to the following limitations:

1. This is a quasi-experimental (observational) study and is therefore subject to the statistical limitations of such a study. Students, for example, were not randomly selected to participate in one of the two "treatment" groups (applied academics versus traditional academics); that decision was made prior to the start of the study by either the students themselves, their parents, their teachers, or their school administrators.
2. Significant correlation coefficients do not necessarily infer causal relationships. If a convincing argument for a relationship is to be made, it must come from a combination of subject matter expertise and common sense, in conjunction with valid methods of statistical inference (Johnson & Wichern, 1992, p. 341).
3. This study was limited to students enrolled in applied academics and corresponding traditional courses at nine public high schools within Iowa.
4. This study was limited to students enrolled in grades 9 through 12 during the 1995-96 school year at the aforementioned high schools within Iowa.
5. This study was limited to the following applied academics courses:
   - Applied Math I and II
   - Principles of Technology I
   - Applied Communications

6. This study was limited to the following traditional academics courses:

   - Algebra I, Algebra II, Geometry, or Trigonometry
   - Physics
   - Basic Communications, Composition, or Composition and Literature

7. This study was limited to the three Work Keys assessment tests that purported to measure skills developed specifically in the applied and traditional courses under investigation. These Work Keys assessment tests are:

   - Applied Mathematics
   - Applied Technology
   - Reading for Information

Definition of Terms

Applied academics: A specific group of courses developed by the Center for Occupational Development (CORD), the Agency for Instructional Technology (AIT), or both. The courses are Principles of Technology, Applied Biology/Chemistry, Applied Math, and Applied Communications. The curricula are written at an 8th grade reading level, incorporate contextual (real world) examples and exercises, and are targeted to the middle 50% of the high school student population.

Concomitant variable: An independent quantitative variable added to a study in which covariance analysis is used. It is important that these variables be observed before the study, or at the very least not influenced by the treatments in any way; else the results could be misleading (Neter, Wasserman, & Kutner, 1990, p. 862).

CORD: an abbreviation for the Center for Occupational Research and Development; an organization in Waco, Texas responsible for a number of applied academics instructional materials.

Employability skills: "... the practical skills that individuals need to obtain and keep jobs" (Boesel & McFarland, 1994, p. 8). These skills can range from technical competence in specific areas, such as mathematics or writing, to interpersonal skills, such as ability to work in a team.

External validity: "... the generalizability or representativeness of the experimental findings. ... What relevance do the findings concerning the effect of X have beyond the confines of the experiment? To what subject populations, settings, experimental variables, and measurement variables can these findings be generalized" (Isaac & Michael, 1990, p. 62)?

Internal validity: The absence of extraneous variables that may impact the dependent variable under investigation and lead one to (erroneously) believe that the independent variables used in the study produced the change in the dependent variable (Isaac & Michael, 1990, p. 60).

Iowa Business Council (IBC): A nonprofit, politically independent group composed of nineteen members (generally presidents or board chairpersons) from major Iowa employers. Together, these companies account for seven percent of the Iowa workforce. IBC's self-described objective is to "develop, through objective research and informed discussion, findings and recommendations for public and private policy that will contribute to preserving and strengthening the quality of life in Iowa through job creation and economic growth" (ACT Center for Education and Work, 1995).

ITED: An abbreviation for the Iowa Tests of Educational Development.

Observational study: A study in which the investigator lacks the power to randomly assign subjects to treatment groups. The investigator is restricted to the choice of observations that are collected and analyzed (Snedecor & Cochran, 1989, p. 14).

Quasi-experimental study: A study in which one attempts to approximate the conditions of a true experimental study, but is unable to control all relevant variables (Isaac & Michael, 1990, p. 42). In many cases this takes the form of being unable to randomly assign subjects to treatment groups.

SCANS: An abbreviation for The Secretary's Commission on Achieving Necessary Skills. This abbreviation is also sometimes used to refer to a June 1991 report by the Commission entitled What Work Requires of Schools: A SCANS Report for America 2000.

VoTech: An abbreviation for Vocational / Technical

Work Keys System: The system developed by the American College Testing's (ACT) Center for Education and Work. The system is composed of four components: job profiling, assessments of personal skill levels, curricula development support materials, and customized reporting capabilities. The goal of the system has been to provide an employability skills metric.

Work Keys Tests: A series of Guttman-based tests designed to assess personal skill levels in key areas associated with employability skills. There are currently eight tests: (a) Applied Mathematics, (b) Applied Technology, (c) Listening, (d) Locating Information, (e) Observation, (f) Reading for Information, (g) Teamwork, and (h) Writing. Only the objective paper-and-pencil tests--Applied Mathematics, Applied Technology, and Reading for Information--were used during this investigation.

Organization of this Study

Chapter 1. Introduction, contains a brief overview of driving forces and circumstances that had an impact on the field of education; eventually leading up to current efforts in applied academics. First, the founding principles of education are briefly reviewed, followed by examples of beliefs, forces, and circumstances that may have guided the evolution of applied academics. A description of the features one currently expects to find in applied academics curricula and an outline of key elements of the Iowa Department of Education

RFP appear just before the statement of the problem. The purpose of the study, the research questions, and the assumptions and delimitations of the study follow the problem statement. The Introduction concludes with a section containing the definitions of terms and an overview of the organization of this study.

Chapter 2. Literature Review, includes a section that gives a historical perspective on VoTech education and the need for an evaluation component in this kind of education. Other sections in this chapter cover applied academics research, and statistical methods used in educational research. A summary of the findings completes this chapter.

Chapter 3. Methodology, covers the research approach and design; the population and selection of the sample; the Work Keys instruments used in the analysis; the data collection and analysis procedures; and finally, the assumptions and limitations of the methodology used.

Chapter 4. Results, describes the sample and variables data used in the study; it also contains sections on the results of exploratory data analysis and data analysis for statistical inference; a section discussing the outcomes of the statistical tests; and finally, a discussion segment.

Chapter 5. Conclusions, summarizes the results of the study, discusses implications of the study for Iowa, and provides suggestions for future research.

CHAPTER 2. LITERATURE REVIEW

A study evaluating the effectiveness of applied academics requires that information be available regarding proposed analysis methods, along with a summary of previous investigative efforts. The decisions regarding the statistical methods to be used are of particular concern since methods recommended for the analysis of this kind of nested data (students within classrooms within schools) have evolved over the last decade. For this reason, a statistical methods section was included along with the review of the applied academics work.

One further note: The word "effectiveness" could perhaps be better defined; as used in this context, effectiveness indicates how well the applied academics component meets its goal of preparing students for the workplace and continuing education.

Organization of this Chapter

The review of literature will be divided into three sections followed by a brief summary. The sections include:

1. Historical Background
2. Applied Academics Research
3. Statistical Methods used in Educational Research

Historical Background

Although major influences affecting the development of industrial and vocational education can be traced back centuries--among others, the publication of Emile by Rousseau in 1762, Pestalozzi's and Fellenberg's schools during the early 1800s, and the Swedish sloyds (mid-1800s) --the vocational education movement in the United States is said to have officially begun in 1906 with the report of the Douglas Commission to the Massachusetts Legislature (Bennett, 1937, p. 507). Bennett references the Report of the Commission on Industrial and Technical Education, State of Massachusetts, 1906, when reporting that the Commission was "to investigate the needs for education in the different grades of skill and responsibility in the various industries of the Commonwealth" (p. 513) and to "consider what new forms of educational effort may be advisable" (p. 513).

Although not explicitly stated above, the Commonwealth of Massachusetts must have had some type of evaluative component in mind. If "needs for education" in various industries are identified and "new forms of educational efforts" are implemented to meet those needs, one would expect some kind of evaluation to be implemented to gauge success in meeting the stated objective. Evaluation of these kinds of educational efforts was certainly nothing new, particularly in Massachusetts. According to Madaus, Scriven, & Stufflebeam, (1983, p. 5, as cited in Hall, 1989, p. 21), there was an attempt to measure the performance of educational programs in Boston in 1845. Hall (1989) mentions that they considered this event to be of particular importance because it "began a long tradition of using pupil test scores as a principal source of data to evaluate the effectiveness of a school or instructional program" (pp. 21-22).

The modern day goal of the Iowa Business Council as described in Making the Grade (ACT, 1995) sounds remarkably similar to the 1906 Douglas Commission goal mentioned previously:

> We [the Iowa Business Council] identified a wide variety of 25 entry level jobs (all above minimum wage) in our companies. Employees in those jobs identified the skills and skill levels needed to perform their jobs. ACT then provided us with a national sampling of high school seniors test results, clearly indicating significant gaps in students' competency to perform well .... .
>
> The findings of the project showed that if Iowa high school graduates--both college and non-college bound--are to obtain a higher degree of employment success in our companies, the curriculum in our schools must become broader and more skills-oriented, and more students must achieve at higher levels in these areas. (p. 3)

In between these two examples were numerous legislative actions designed to further the development of Vocational and Technical Education in this country. These actions included:

- The Smith-Hughes Act in 1917 designed to provide continuing appropriations for vocational education in agriculture, trades and industry, homemaking, and teacher training.
- The George-Reed, -Ellzey, -Deen, -Barden Acts in 1929, 1934, 1937, and 1946 respectively, to provide funds for home economics, agricultural education, and distributive education.

- The Vocational Education Act of 1963 and The Vocational Education Amendments of 1968 and 1976
- The Job Training Partnership Act of 1982
- The Carl D. Perkins Vocational Education Act of 1984

A brief discussion of the above Acts is contained in Administration of Vocational Education (Wenrich, Wenrich, and Galloway, 1988, pp. 28-35). Of particular importance however is the introduction of evaluation components in the later Acts. Wenrich et al. (1988) mention that, "The authors of the Vocational Education Act of 1963, recognizing the need for flexibility in a rapidly changing society and the difficulties of reorienting institutions to keep pace with new demands, built into the Act an evaluation system" (p. 28). Following the Perkins Act of 1984, the Perkins Act of 1990 provided certain vocational education programs with federal assistance through June 30, 1996. According to the AVA guide to the Perkins Act (AVA, 1990) strong evaluation components were a part of this legislation. Each state must:

> ... develop and implement a statewide system of core standards and measures of performance for secondary, postsecondary, and adult vocational education programs. ... Annually, each recipient must evaluate the effectiveness of the programs conducted with Perkins funds, based on the core standards and measures of performance. (p. 12)

The Tech-Prep Education Program is funded under Title III of the 1990 Perkins Act. Under Title IV of the 1990 Perkins Act, "The Secretary is authorized to provide funding to a wide range of educational institutions forming consortia to develop, implement and operate programs using different models of curricula which integrate vocational and academic learning" (page 14). Some of the efforts to design, implement, operate, and evaluate applied academics programs in secondary schools are described below.

Applied Academics Research

A 1995 paper by Wang and Owens covers fourth-year results of a Boeing Company-funded applied academics project. This Evaluation Report indicated:

> ... AM [Applied Mathematics] students in every category scored significantly higher than their peers in traditional mathematics classes. PT [Principles of Technology] students performed as well as their traditional counterparts when the variables for overall GPA [Grade Point Average] and grades in mathematics and science were held constant. Low-achieving AM and PT students tended to demonstrate the greatest gain from the applied academic courses. (abstract)

Major findings reported in the above study included the following:

- The group of Applied Mathematics students had similar overall GPAs, mathematics grades, and educational aspirations as the group of comparison students. One noted difference between groups was that a lower percentage of the Applied Mathematics students reported that they would like to be employed immediately after high school (33 percent versus 44 percent).
- The sampled group of students in Principles of Technology classes had lower overall GPAs, lower overall mathematics and science grades, and lower educational aspirations (measured as the percentage of the group planning to attend a university) than the comparison group.

Wang and Owens (1995) also included sections on student demographics (gender, grade level, and race) and, in both the Applied Mathematics (AM) and Principles of Technology (PT) studies, made an effort to introduce some of these demographic variables in the analyses. For example, they reported that: "By controlling gender, grade level, overall grade in mathematics, and overall GPA, AM students still scored significantly higher than comparison students" (p.16). In the Principles of Technology section, they stated that: "Female students in both groups scored significantly lower than male students in the post-test" (p. 20). They also stated that: "... PT students at grade 9 scored significantly higher than did PT comparison students of the same grade level. No difference was found at grade 10. At grades 11 and 12, the comparison students scored significantly higher than PT students" (p. 20). Finally, Wang and Owens (1995) mention that:

> The results of Factorial ANOVA indicated that the level of algebra used as a covariate had a significant impact on both PT and comparison student's scores on the post-PT test. When overall grades in mathematics and science and the overall GPA were controlled, PT students generally scored higher (in some cases significantly higher) than did the comparison students on PT items in the post-test. (p. 20)

Wang and Owens (1994) presented a paper at the Annual Meeting of the American Educational Research Association titled "A Multiple Approach to Evaluating Applied Academics". In this paper they provide an overview of applied academics and a description of the approaches the Northwest Regional Educational Laboratory has used in evaluating a specific applied academics project undertaken in partnership with the Boeing Corporation. They state that: "Although students' performance on relevant tests is a good indicator of the effectiveness of applied academics, our expectations and assessments should also reflect the different approaches to teaching and learning in applied academics" (p. 8). They outline a series of criteria by which Hull and colleagues would reportedly evaluate an applied academic curriculum. An applied academics curriculum would be considered to be a success if:

- Students are able to transfer knowledge from academic content to vocational applications and from school to the workplace.
- Students are not afraid to take academic subjects such as mathematics and science.
- Students display more interest, motivation, and understanding of the value of the subject and of school in general than they did in classes taught by traditional methods.
- The applied course is as challenging as the traditional "college-prep" course on the same subject--not low level or watered down.
- The student population that has traditionally done poorly in academic subjects displays improved performance.
- Applied courses receive the same recognition and acceptance from universities and colleges as do the traditional courses with the same content. (as cited in Wang & Owens, 1994, p. 8)

The Wang and Owens (1994) paper is a discussion of approaches, not results of their use.

Dugger and Johnson (1992) described a summative evaluation of the Principles of Technology (PT) program. The study compared student achievement regarding basic physics concepts as measured by a 120 question PT instrument. Two treatment groups and a control group were involved in the study. One treatment group consisted of students enrolled in first year PT classes; the second treatment group consisted of students enrolled in high school physics classes. Fifteen Iowa high schools participated in the study. These schools were chosen because they had offered both PT and physics as part of the regular curriculum for at least two years. Students in the control group were selected from a pool of students enrolled in neither PT or physics classes. The control group students were chosen to ensure a gender ratio and ITED (Iowa Tests of Educational Development) score distribution similar to that of the PT treatment group. The PT instrument was administered twice, at the beginning of the academic year and at the end of the academic year. The paper reports pre- and post-test means, standard deviations, and sample sizes of the three groups. The t-statistics comparing difference of the pre- and post-test means of the three groups are also included. The differences between the means for the PT and the Physics Treatment groups are noted as significant at the 0.01 level. Based on the study, the authors conclude that: "Although never intended to replace Physics, the Principles of Technology first year course does a significantly better job increasing student achievement regarding basic physics concepts as defined by the Principles of Technology program" (p. 25). This statement is followed however with a warning: "One must exercise caution in drawing inferences regarding the two programs since physics also is responsible for covering higher level concepts that are not considered basic and may be considered non-intuitive" (p. 25).

Dugger and Meier (1994) used the same methodology described in the above Dugger and Johnson (1992) study to evaluate second-year PT and physics student achievement. This study again compared student achievement regarding basic physics concepts as measured by a 120 question PT instrument. Note: Although similar in length, the PT instruments used in this and the previous 1992 study covered different curriculum objectives.

Two treatment groups and a control group were involved in the 1994 Dugger and Meier study. The first treatment group consisted of students enrolled in second year PT classes; the second treatment group consisted of students enrolled in comparable high school physics classes. Three Iowa high schools participated in the study. These schools were chosen because they had offered both PT and physics as part of the regular curriculum for at least three years. Students in the control group were selected from a pool of students enrolled in neither PT nor physics classes. Control group students were chosen to ensure a gender ratio and ITED (Iowa Tests of Educational Development) score distribution similar to that of the PT treatment group. The PT instrument was administered twice, at the beginning of the academic year and again at the end of the academic year. The paper reports pre- and post-test means, standard deviations, and sample sizes of the three groups. The t-statistics comparing difference of the pre- and post-test means of the three groups are also included. The differences between the means for the PT and the Physics treatment groups are noted as significant at the 0.01 level. Two one-way analysis of variance

(ANOVA) tables were also included in the paper; one dealing with pretest results, the second covering post-test results. The between-treatment F-statistics were significant at the 0.01 level in both cases. The authors conclude that: "Exposure to traditional physics does produce significant achievement gains on a second-year Principles of Technology achievement instrument. Even greater significant gains occur if these students are exposed to a second year Principles of Technology course" (p. 11).

Grieve (1990) completed an evaluation of the Applied Academics Options Program for Business students of Greene County Career Center in Ohio. The objectives measured were program enrollment, job placement of graduates, and postsecondary education of graduates. She concluded that the Options Program is meeting its goals in two of the three areas, job placement and postsecondary education, and "may be meeting the third goal [increased program enrollment] to a certain degree" (p. 23).

The Washington State Supervisor of Business Education in 1989 provided a brief report on the administrative steps for implementing applied academics in a 1994 National Business Education Yearbook article (Shaw, 1994). The report was based on the experiences of the group charged with developing an applied communications implementation plan. According to Shaw, cooperation of the state academic and vocational staffs is the key element in any applied academics implementation project. The article provides some fairly specific information regarding the implementation procedures and guidelines; including a detailed agenda for the 30-hour Applied Communication Training Session. Two other items of note were mentioned in the article:

1. The state of Washington contracted with AIT to provide courseware for the Applied Communications modules and only those districts approved for implementation were allowed to purchase materials from AIT.
2. The entire applied academics project in the Washington received a boost from a major industrial partner located in the state. Shaw (1994) writes,

> Because the citizens of the state of Washington are very generous supporters of education, and business requires an educated workforce, the Boeing Company pledged its support to the applied academic programs. Boeing's Foundation in 1990 established an unprecedented partnership with schools. Through the foundation, over two million dollars was [sic] contributed in grants for equipment, in-service of instructional staff, materials, and instructional industry internships. Boeing felt strongly that these programs were designed to make learning meaningful within the context of work and that the curricula stressed application of subject matter. There was strong consensus that the programs delivered the same concepts found in academic disciplines and continue on to deliver the subject matter relative to real-life tasks. (p. 30)

Statistical Methods

This study investigates the differential effectiveness of two instructional methods. There is a concern that results of the data analysis may be confounded by the effect of group differences in independent variables. Two primary concerns have surfaced during research regarding statistical techniques used in this type of educational analysis: one deals with the use of covariates; the other with the hierarchical nature of the data and its impact on the unit of analysis.

Covariates

A common method of controlling extraneous variance in such situations is by analysis of covariance techniques (ANCOVA). The use of such a technique, however, is not without its problems. Pedhazur (1982) provides an excellent general discussion of the use of ANCOVA in the social sciences. There are several segments of that discussion that are extremely relevant to this investigation.

> Other examples of the use of ANCOVA in quasi-experimental designs are encountered in settings in which, because of administrative or other considerations, subjects cannot be assigned randomly to treatments. Instead, the treatments are administered to intact groups. This happens very often in educational settings. For example, suppose again that one is studying the effects of different teaching methods on achievement. The school in which the study is conducted does not permit random assignment of pupils to treatments, but insists that intact classes be used. The researcher suspects, or knows, that classes differ in mental ability. Under such conditions, it is possible that the classes higher in mental ability will perform better regardless of the teaching method to which they are assigned. This may therefore lead to the erroneous conclusion that a given method is superior, when its apparent superiority is due to the mental ability of the subjects assigned to it. To avoid such a blunder, the researcher attempts to "equalize" the groups on mental ability by the use of ANCOVA. In other words, an attempt is made to take into account, or "adjust" for initial group differences in mental ability. The same reasoning is extended to more than one variable. Thus if the researcher believes that the groups differ in motivational as well as mental ability, an "adjustment" is made for initial group differences on both variables; that is, both are used as covariates. (p. 495)

The above quotation seems to speak directly to the controlling factors surrounding this investigation; particularly if one makes the assumptions that a grade point average or a composite score on the Iowa Test of Educational Development may be used as a covariate controlling for supposed differences in mental ability, or that a student's grade level may be used as a covariate to control for certain other factors in a student's educational or emotional makeup.

It must be noted that there is appreciable concern regarding the application of ANCOVA to these kinds of situations. Pedhazur (1982) states that: "Unfortunately, the applications of ANCOVA in quasi-experimental and nonexperimental research are by and large not valid" (p. 496). He goes on to report that:

> Following sound principles of research design, ANCOVA may serve a very useful purpose of control in experimental research.
>
> The situation is radically different (some say hopeless) when ANCOVA is used in quasi-experimental or nonexperimental research for the purpose of "equating" intact groups. The logical and statistical problems that arise in such situations are so serious that some authors have argued that ANCOVA should not be used at all. (p. 520)

The areas of primary concern identified by Pedhazur (1982, pp. 521-525) include Specification Errors, Extrapolation Errors, Differential Growth, Nonlinearity, and Measurement Errors. Each of these areas will be discussed using Pedhazur's framework in the context of this investigation. This should allow the reader a more complete understanding of the concerns of methodologists regarding the use of ANCOVA in quasi-experimental designs.

Specification Errors may result when some of the variables on which groups differ are left uncontrolled in a quasi-experimental ANCOVA design. Failure to control certain variables may result in attributing differences in employability skills to instructional methods, when in fact one or more of the covariates are correlated with variables not included in the equation but which are nevertheless related to the dependent variable. Specification errors will lead to biased estimation of parameters, which in turn leads to overadjustment or underadjustment of treatment means. It is relatively easy to come up with variables that are not controlled in this investigation, but may affect "employability skills." The list could include, but is certainly not limited to the socioeconomic status of the student's family, student work experience, teacher-to-pupil ratio, district expenditures for course materials and equipment, etc. Failure to control for initial differences among intact groups could conceivably lead to an erroneous conclusion that an instructional method is harmful when the opposite is true.

Extrapolation Errors are most prevalent when there is little or no overlap between the distributions of the treatment groups. If ITED scores were used as a covariate and the range of ITED scores for all students in the applied academics courses fell between 30 and 50 while the range of ITED scores for all students in the traditional academics courses fell between 70 and 90 there would be cause for concern regarding extrapolation error.

Differential Growth errors occur when one erroneously assumes that the rates of growth of the individuals in the intact groups are the same. In any learning environment, some students pick up new skills more quickly than others. If a majority of "slow learners" happen to fall in one treatment group, while a majority of the "fast learners" fall in the other, it is possible to conclude that one instructional method is superior to another when that is not the case.

Nonlinearity errors arise when one uses linear regression analysis methods if the regression of the dependent variable on the covariate is linear when it is not.

Measurement Errors in the independent variables or covariates may lead to overestimation or underestimation of the regression coefficients. It is safe to assume in this investigation that there are errors of measurement in one or more of the covariates. Psychological constructs, such as intelligence and academic motivation, are being approximated by student ITED composite scores or student GPA. Crocker and Algina (1986) note some problems commonly encountered with measurements of psychological constructs:

> 1. No single way of defining a psychological construct is universally accepted.
> 2. Psychological measurements are based on samples of behavior.
> 3. *Sampling of behavior results in errors of measurement* [author's emphasis].
> 4. The units of measurement are not well defined. (p. 13)

Pedhazur (1982) provides some additional comments regarding not only random errors of measurement, but also other types of errors of measurement. He states:

> The effects of other types of errors are even more complicated, and little is known about them. But even if one were to consider the effects of random errors only, it is clear that they may lead to serious misinterpretations in ANCOVA. What, then, is the remedy? Unfortunately, there is no consensus among social scientists about the appropriate corrective measures in ANCOVA with fallible covariates. ...
>
> It is not possible to discuss here the different proposed solutions to deal with fallible covariates without having to go into complex issues regarding measurement models. The purpose of the discussion here was only to alert you to the problem in the hope that you will reach two

> obvious conclusions: (1) that efforts should be directed to construct measures of the covariates that have very high reliabilities, and (2) that ignoring the problem, as is unfortunately done in most applications of ANCOVA, will not make it disappear. (p. 524)

The preceding ANCOVA discussion serves to highlight the considerations of which one must be aware when engaged in analysis of a quasi-experimental design. One should not, however, assume that based on the previous discussion quasi-experimental designs have no place in research. Pedhazur (1982) remarks that:

> ..., the discussion was not intended to convey the idea that research other than experimental hold no promise for the social sciences, and should therefore be avoided. On the contrary, there are many good reasons for choosing to study certain phenomena in quasi-experimental or nonexperimental settings. And because of ethical considerations, economic or societal constraints, this type of research may be the only feasible one in various areas. But the conduct of such research, indeed all scientific research, requires sound theoretical thinking, constant vigilance, and a through understanding of the potential and limitations of the methods being used. (p. 525)

One other comment regarding quasi-experimental and nonexperimental studies is particularly relevant to this investigation. Weisberg (1979) expresses his opinion that:

> Given the state of current methodology in the social sciences, the full potential of such studies will not be realized until more appropriate methods, suited to deal with the unique problems they pose, are developed. Attempting to develop such designs ought to be a top priority of evaluation methodologists. Until we have tried to develop alternatives based not on "approximations" to randomization, we should be cautious in discounting the value of uncontrolled studies. While statistical adjustments are certainly problematic, the potential contribution of uncontrolled studies has not really been tested. (p. 1163)

Unit of Analysis

The unit of analysis to be used in this investigation is not entirely clear cut. Pedhazur (1982) in discussing the question of appropriate units of analysis in social and behavioral research says: "In educational research, for example, should the unit of analysis be the individual student, the class, the school, the school district, the state" (p. 526)? He elaborates on this point using a fictitious example described below:

> .... Assume that classes are randomly assigned to one of two instructional methods and that it is desired to study whether there is an interaction between the methods and some aptitude of the students. This, then, is the simplest example of a study in which the researcher is faced with the dilemma of whether to use the student or the class as the unit of analysis. Traditionally, most researchers have focused on the student, although some have chosen to focus on the class instead.
>
> [Several authors cited by Pehazur on page 545] argue, however, that the choice of one level to the exclusion of the other may result in either masking certain effects or in indicating effects where none exist. This is because certain processes operate on the group level and others operate on the individual within a specific group. Thus, for example, Cronbach and Webb (1975) reason that a high mean aptitude of a class may lead a teacher to crowd more material into the course, thereby leading to either greater or lesser achievement for the class as a whole. Treatments may also have "comparative effects within a group" (Cronbach & Webb, 1975, p. 717). If, for example, "one method provides special opportunities or rewards for whoever is ablest within a class, the experience of a student with an IQ of 110 depends on whether the mean of his class is 100 or 120" (Cronbach & Webb, 1975, p. 717). Accordingly, Cronbach (1976) proposes, among other things, that between-class regression coefficients be studied for the purpose of detecting processes that operate on classes as units, and that within-classes regression coefficients be studied for the purpose of detecting processes that act on individuals as units within classes. ...
>
> The important role of multilevel analyses when individuals are nested within groups, and groups are nested within larger units (e.g., instructional methods) cannot be overemphasized. (p. 546)

Iversen (1991) supports the notion of multilevel analysis: "The context of a group is believed to affect the actions and attitudes of the individuals who belong to the group" (p. 3). He goes on to explain that:

> Contextual analysis is possible when we have data on individuals as well as data on the groups to which they belong. Typically, we have a dependent variable measured on the individuals, and we want to study the effects of characteristics of the individuals themselves as well as the characteristics of the groups to which they belong. Most of the time we want to find out what the form is of the effect of the individual and the group variables. We want to find out whether the effects are positive or negative, and we want to find out whether the effects are linear or nonlinear. We also are interested in whether the individual and group characteristics act together to produce interaction effects. (p. 3)

In Bryk and Raudenbush (1992), the Series Editor, Jan de Leeuw also discusses the unit of analysis issue in the context of hierarchical data. He states that:

> The next step, after realizing how important hierarchical data are, is to think of ways in which statistical techniques should take this hierarchical structure into account. There are two obvious procedures that have been somewhat discredited. The first is to disaggregate all higher order variables to the individual level. Teacher, class, and school characteristics are all assigned to the individual, and the analysis is done on the individual level. The problem with this approach is that if we know that students are in the same class, then we also know that they have the same value on each of the class variables. Thus we cannot use the assumption of independence of observations that is basic for the classical statistical techniques. The alternative is to aggregate the individual-level variables to the higher level and do the analysis on the higher level. Thus we aggregate student characteristics over classes and do a class analysis, perhaps weighted with class size. The main problem here is that we throw away all the within-group information, which may be as much as 80% or 90% of the total variation before we start the analysis. As a consequence, relations between aggregated variables are often much stronger, and they can be very different from the relation between the nonaggregate variables. Thus we waste information, and we distort interpretation if we

> try to interpret the aggregate analysis on the individual level. Thus aggregating and disaggregating are both unsatisfactory. (p. xiv)

Jan de Leeuw goes on to say that:

> Hierarchical linear models, or multilevel models, are certainly not a solution to all the data analysis problems of the social sciences. ... Nevertheless, technically they are a big step ahead of the aggregation and disaggregation methods, mainly because they are statistically correct and do not waste information. (p. xv)

Bryk and Raudenbush (1992) themselves note that:

> Educational research is often especially challenging because studies of student growth often involve a doubly nested structure of repeated observations within individuals, who are in turn nested within organizational settings. Research on instruction, for example, focuses on the interactions between students and a teacher around specific curricular materials. These interactions usually occur within a classroom setting and are bounded within a single academic year. The research problem has three foci: the individual growth of students over the course of the academic year (or segment of a year), the effects of personal characteristics and individual educational experiences on student learning, and how these relations are in turn influenced by classroom organization and the specific behavior and characteristics of the teacher. Correspondingly, the data have a three-level hierarchical structure. The Level-1 units are the repeated observations over time, which are nested within the Level-2 units of persons, who in turn are nested within the Level-3 units of classrooms or schools. (p. 2)

Bryk and Raudenbush (1992) also provide numerous examples of the use of Hierarchical Linear Models (HLM), including a specific Three-Level Model consisting of students (Level-1) nested within classroom (Level-2) nested within schools (Level-3).

One other technique used in data analysis deserves mention; the use of gain scores as a dependent variable. According to Cronbach and Snow (1977):

> Their use is particularly common in studies of school learning, where data are collected only before and after instruction. Although the simple difference between pretest and posttest seems to measure learning rate, such scores are likely to produce misleading results. One has to assume that test scores form a meaningful interval scale. This common assumption is rarely troublesome in ordinary measurement. But in dealing with gains the assumption of equal intervals is critical and it almost never can be defended. Also, the gain score is likely to be highly unreliable, since it combines the errors of two fallible measures. Errors introduce systematic biases; for example, the person who happens to be unlucky in his guesses on the pretest will for that reason show what seems to be a fine improvement.
>
> Lord (1956, 1958, 1963), McNemar (1958), and Cronbach and Furby (1970) have suggested how true gain can be estimated so as to overcome many kinds of bias, though no estimate of a gain score can overcome the problem of scale intervals. Cronbach and Furby, however, after considering how to improve estimates, argued that estimates of change scores should rarely or never be employed. For describing and testing treatment effects they advocated the use of the observed posttest score, or some composite of posttest scores, as the dependent variable. (p. 73)

Crocker and Algina (1986) would also seem to discourage the use of gain scores in most instances. They state that:

> Difference scores, or gain scores between two testings, are usually less reliable than scores on either single testing occasion when errors of measurement are uncorrelated. Reliable difference scores can be obtained only by using tests which are highly reliable at each occasion and which have low correlation between them. (p. 153)

Summary

There has long been an interest in evaluating the effectiveness of various types of instructional methods. To aid in an understanding of the pertinent issues of this research, a literature review was performed covering three primary areas for this chapter. The first section included a review of some of the more important historical milestones in the development of applied academics; the second provided an overview of current published findings in applied academics research; and the final section addresses the question of what statistical techniques are suitable for this investigation.

Most of published studies indicate favorable results using applied academics curricula although some of the findings are anecdotal and, those that are not, do not appear to take into account the hierarchical nature of the data. Hierarchical models and the software needed to allow widespread use of such models are still relatively new to educational research and previous researchers may have been unaware of these techniques.

CHAPTER 3. METHODOLOGY

Overview

This dissertation is the product of a focus on one element of a much broader study. In August of 1995, the Iowa Department of Education sent out a Request for Proposals (RFP) regarding an evaluative study of applied academics in Iowa. This was to be a 10 month project with a start date of October 1, 1995. A proposal submitted by a team from the Department of Industrial Education and Technology (IEdT) and the Research Institute for Studies in Education (RISE) at Iowa State University was ultimately accepted. Dr. Jan Sweeney, Dr. Mandi Lively, and Mari Kemis of RISE were involved in the initial aspects of the project, however during the course of the project it was turned completely over to investigators operating out of the IEdT department. Team members, besides the author, who were actively involved in the project during the execution of the contract included Dr. John Dugger, Dr. Oscar Lenning, and Ms. Andrea Wright.

Research Approach

A quasi-experimental method of research was used in this investigation. Ideally, one would prefer to use random selection methods to identify both schools and students within those schools for participation; however, true random selection was not possible. Although there are many difficulties inherent in true experimental research using human subjects, it was not possible in this case to even attempt a true experimental study; classes had already started when the study began and random assignment of students to classes was not an option. This coupled with the time and funding restrictions built-in to the Department of Education contract made the choice of a quasi-experimental method the best available option.

Research Design

As stated in the RFP, the intent of the study was to evaluate the effectiveness of the applied academics component of Iowa's Tech Prep effort in meeting the employability skills needs of Iowa employers. In particular, the Iowa Department of Education requested a research program to investigate the following specific questions:

1. Is there a difference in student academic achievement for students who have completed applied academic courses in contrast to those who have completed comparable traditional academic courses?
2. Are the employability skills of students improved by their completion of applied academic courses in contrast to the employability skills of students who have completed comparable traditional academic courses?

In order to quantitatively estimate the difference in academic performance between students who have completed applied courses and students who have completed comparable traditional courses (i.e., the first research issue), the following questions and hypotheses were devised:

1.1 Is there a statistically significant difference between the paired course mean high school GPAs of students who have completed applied courses versus those who have completed comparable traditional courses?

Null hypothesis: There is no statistically significant difference between the paired course mean high school GPAs of the two groups.

1.2 Is there a statistically significant difference between the paired course mean composite ITED scores of students who have completed applied courses versus those who have completed comparable traditional courses?

Null hypothesis: There is no statistically significant difference between the paired course mean composite ITED scores of the two groups.

To quantitatively estimate the relationships between curricula and student's employability skills as per the second research question, the independent, dependent, and classificatory variables had to be precisely defined and the above questions rephrased in such a manner as to permit statistical inferences consistent with the assumptions and limitations of the study. The "treatments" under investigation in this study were applied curricula and traditional curricula. The dependent variables included:

- Work Keys Applied Mathematics test score,
- Work Keys Applied Technology test score, and
- Work Keys Reading for Information test score.

The above dependent variables were used as measures of employability skills. It was not the intent of this study to question the reliability and validity of Work Keys test scores as measures of employability skills. The assumption was made that they do serve as valid and reliable measures of employability skills. The following "Instrumentation" section contains information supplied by the test developers as to test reliability and validity so that readers may evaluate the appropriateness of this assumption for themselves.

The classificatory variables that were considered for use in the investigation included:

- Student grade level;
- Student Iowa percentile rank for the Iowa Tests of Educational Development (ITED) composite score;
- Student Grade Point Average (GPA);
- Student gender;
- Type of course (applied or traditional); and
- Whether a course was relevant to the Work Keys test taken by the student.

The questions and hypotheses devised to enable the use of statistical inference techniques for the second research question included:

2.1 Is there a statistically significant difference in the raw class mean Work Keys scores between students who have completed applied courses and students who have completed traditional courses?

Null hypothesis: There is no statistically significant difference in the raw class mean Work Keys scores between groups of students who have completed applied courses and students who have completed traditional courses.

2.2 Are there statistically significant correlations between Work Keys raw class mean scores and the following concomitant variables: (a) student grade level, (b) student grade point average (GPA), (c) student composite score on the Iowa Tests of Educational Development (ITED), or (d) student gender?

Null hypothesis: There are no statistically significant correlations between Work Keys raw class mean scores and the concomitant variables.

2.3 Is there a statistically significant difference in the adjusted class mean Work Keys scores between students who have completed applied courses and students who have completed traditional courses? The word "adjusted" indicates that the raw Work Keys scores have had the mean effects of any known significant concomitant variables removed.

Null hypothesis: There is no statistically significant difference in the adjusted class mean Work Keys scores between groups of students who have completed applied courses and students who have completed traditional courses.

Population and Selection of Sample

The population for the study consisted of high school students, grades 9 through 12, enrolled in one of the nine Iowa public high schools listed below:

1. Cascade Junior-Senior High School
   505 Johnson Street NW
   Cascade, Iowa 52033

2. Le Mars High School
   921 Third Avenue SW
   Le Mars, Iowa 51031

3. Moc-Floyd Valley High School
   615 Eighth Street SE
   Orange City, Iowa 51041

4. Sigourney Junior-Senior High School
   RR 2
   Sigourney, Iowa 52591

5. Sioux City North High School
   4200 Cheyenne
   Sioux City, Iowa 51104

6. South O'Brien High School
   307 West Groesbeck
   Paullina, Iowa 51046

7. Spencer High School
   800 East 3rd Street
   Spencer, Iowa 51301

8. West Marshall High School
   State Center, Iowa 50247

9. Western Dubuque High School
   5th Avenue West
   Epworth, Iowa 52045

The sample for the study includes 1,321 students, split relatively evenly between applied and traditional courses. The students were enrolled in one of the above nine high schools during the 1995-1996 school year.

Instrumentation

Three Work Keys assessment test scores serve as a measure of employability skills. The three tests used in this study include ***Reading for Information***, ***Applied Mathematics***, and ***Applied Technology***. According to the developer, American College Testing (ACT, 1994), Work Keys is "a national system for documenting and improving workplace skills" (p. 1). American College Testing (1994) provides the following summary of the Work Keys system:

> The Work Keys System has as its basis a metric, or measurement scale, that can be used to compare an individual's employability skills with the requirements of a particular job .... Work Keys provides a universal metric that translates skill requirements for individual jobs into "levels" of proficiency. Such a metric makes it possible for schools to determine how to prepare students more completely for the workplace. It also helps businesses determine the qualifications of potential employees, as well as design job-training programs that will help current employees meet the demands of their jobs. (pp. 1-4)

The development of the Work Keys System was not solely the effort of ACT. ACT used advisory panels of business people and educators to assist in the design and review of plans and materials. These panels were also instrumental in identifying and recruiting examinees for the prototype and field-test phases of assessment development. They were also involved in identifying and recruiting content-qualified and minority individuals to conduct content and fairness reviews of the assessments. Members from the educators group included individuals from California, Illinois, Iowa, Michigan, Ohio, Oregon, Tennessee, and Wisconsin (ACT, 1994, 1996).

ACT (1994) provides a detailed description of the assessment development process in their Psychometric Handbook draft. A brief, paraphrased, description of their process is included below.

The development process consists of four phases: developing test specifications, prototyping, pretesting, and constructing operational forms. First the "blueprints", or test specifications, for specific tests were developed. The blueprints provide test specialists with an overall structure to tell them what skills the assessment is going to measure and how the items will increase in complexity as the skill level increases. A panel of business people, educators, and ACT staff then reviewed the test specifications and recommended any changes they think would be valuable. Once the foundation of the specifications was set, the test specialists began writing items for the prototype tests. In the prototyping phase, test specialists and item writers developed a small number of items corresponding to each level of difficulty. All items were written to the specifications developed by ACT and the advisory panels and, once written, were edited by ACT staff. The prototype tests were administered to small groups of students and employees to determine whether the test specifications were functioning properly.

After the prototyping phase, the item writers produced a large number of test items that the ACT staff edited to meet content, cognitive, and format standards. The editing process included content and fairness reviews by the content-qualified and minority reviewers respectively. The items were then pretested by large enough samples of students and employees to permit evaluation of each test item's psychometric properties (for example, reliability and scalability). Questionable items identified during the pretest phase were reevaluated by ACT test specialists or, if necessary, content reviewers external to ACT.

Finally, the ACT staff constructed operational forms, better known as test booklets, of the specific assessments. Alternate and equivalent forms of the assessment were also developed from the pool of items meeting all content, statistical, and fairness criteria.

The measurement scale chosen for assessments is Guttman-based. As ACT (1994) states in its documentation:

> Establishing a measurement scale appropriate for the development of the Work Keys assessments was especially challenging. This measurement scale had to support not only the evaluation of individuals' skills, but the evaluation of job requirements, as well. Because of this unique duality of a single metric describing both individual skills and job requirements, it was decided that the Work Keys score scale should be **Guttman-based.** (pp. 14-15)

According to Isaac and Michael (1990), Guttman-type scales:

> ... consist of a relatively small set of homogeneous items that are supposedly unidimensional, measuring one, and only one, attribute. Such scales get their name from the cumulative relation between items and the

> total scores of individuals. Items can be ordered in difficulty, complexity, or value-loading (from low to high) so that to answer correctly or approve the last implies success or approval on all the preceding ones; or, to miss or disapprove a middle item implies failure or disapproval on all the subsequent items. When the scale is cumulative and we know a subject's total score, we can predict his answering pattern. (p. 143)

The operational forms of the three Work Keys assessments used in this study as described by ACT (1994) are found below.

> - ***Reading for Information*** - This assessment measures the examinee's skill in reading and understanding work-related reading materials. It is designed to reflect the types of reading matter found in a variety of business settings. The reading passages and questions are based on the actual demands of the workplace. Selections take the form of memos, bulletins, notices, letters, policy manuals, or governmental regulations. [*This assessment scale runs from Level 3, the least complex, to Level 7, the most complex.*]
> - ***Applied Mathematics*** -This assessment measures the examinee's skill in applying mathematical reasoning to work-related problems. The test questions require the examinee to set up and solve the types of problems and do the types of calculations that actually occur in the workplace. This test is designed to be taken with a calculator (supplied either by the examinee or by the agency administering the assessment) because, on the job, the calculator serves as a tool for problem solving. A formula sheet that includes, but is not limited to, all required formulas is provided. [*This assessment scale runs from 3 to 7.*]
> - ***Applied Technology*** - This assessment measures the examinee's skill in solving problems of a technical nature. The content covers the basic principles of mechanics, electricity, fluid dynamics, and thermodynamics as they apply to machines and equipment found in the workplace. Because the assessment is oriented more toward reasoning than toward mathematics, any calculations required to solve a problem can be readily performed by hand. The emphasis is on identifying relevant aspects of problems, analyzing and ordering those aspects, and applying existing materials or methods to new situations. [*This assessment scale runs from 3 to 6.*] (pp. 15-16)

ACT staff also investigated the psychometric properties of the Work Keys assessments. They computed four indices for the assessments: coefficient of reproducibility, minimal marginal reproducibility, percent of improvement, and coefficient of scalability. Although the actual results of the computations were not available in the information provided to this author by ACT, lower threshold figures were given in their documentation for two of these indices. ACT (1994, p. 20) set the acceptable limits for inferring that a scale is "Guttman-like" at 0.90 and 0.60 for the coefficients of reproducibility and scalability respectively.

ACT staff (1994) also provided information regarding classification consistency and reliability of the "current" [1994] operational formats, but warned readers that: "Given the nature of the scale and scoring of the Work Keys selected-response assessments, internal consistency or coefficient alpha is not an appropriate reliability statistic" (pp. 22-23). The then-current assessments yielded the following results.

- ***Reading for Information***: 30 items over 5 levels, alpha = 0.82
- ***Applied Mathematics***: 30 items over 5 levels, alpha = 0.83
- ***Applied Technology***: 32 items over 4 levels, alpha = 0.75

ACT staff preferred another measure over coefficient alpha (ACT, 1994).

> A better sense of the reliability or consistency of the Work Keys assessments is given by the number of misclassified individuals with respect to the highest contiguous level achieved. ... Normally, with the exception of the Teamwork assessment, classification consistency is consistently in the 95-percent or above range. (p. 23)

The size and demographics of the sample reported used by ACT (1994) are described below:

> The sample analyzed for this technical summary contained 14, 584 examinee records. Of these examinees, 72 percent marked their ethnic origin as Caucasian, 11 percent marked it as African American/Black, two percent marked Native American, five percent marked Hispanic (all categories combined), one percent marked Asian, and nine percent either did not mark their ethnic origin or marked the "prefer not to respond" option. The average age of these examinees at the time of testing was 18 years, with a standard deviation of ten years. Fifty-three percent of the sample were in twelfth grade at the time of testing. ... Each examinee took at least two of the Work Keys assessments. (pp. 18-19)

The question of validity is also important to the Work Keys system. According to a group composed of members from the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education (1985), validity is an indication of:

> ... the appropriateness, meaningfulness, and usefulness of the specific inferences made from test scores. Test validation is the process of accumulating evidence to support such inferences. A variety of inferences may be made from scores produced by any given test ... . Validity, however, is a unitary concept. Although evidence may be accumulated in many ways, validity always refers to the degree to which that evidence supports the inferences that are made from the scores. The inferences regarding specific uses of a test are validated, not the test itself. (p. 9)

A 1996 ACT publication subsequent to the Psychometric Handbook entitled Work Keys: Validity Supplement addresses the issue of validity. In this document ACT approaches the question of validity using a three-part model. This model looks at the triad of content validity, criterion-related validity, and construct validity.

Regarding content validation, Crocker and Algina (1986) state that:

At the minimum, content validation entails the following steps:
1. Defining the performance domain of interest
2. Selecting a panel of qualified experts in the content domain
3. Providing a structured framework for the process of matching items to the performance domain
4. Collecting and summarizing the data from the matching process (p. 218)

According to ACT (1996) the Work Keys system tests were designed to meet the following criteria.

1. The way in which the generic skill is assessed is generally congruent with the way the skill is used in the workplace.
2. The lowest level assessed is at approximately the level for which an employer would be interested in setting a standard.
3. The highest level assessed is at approximately the level beyond which specialized training would be required.
4. The steps between the lowest and highest levels are large enough to be distinguished and small enough to have practical value in documenting workplace skills.
5. The assessments are sufficiently reliable for high-stakes decision making.
6. The assessments can be validated against empirical criteria.
7. The assessments are feasible with respect to administration time and complexity, as well as cost. (pp. 6-7)

ACT did use panels of qualified content domain experts in the test development process as was previously explained. The development method included both advisory panels and two types of reviewers of the proposed material (see Table 3.1). The advisory panels were composed of business people and educators knowledgeable in the areas in question; the reviewers were outside experts knowledgeable in content and fairness issues.

Table 3.1. Number of reviewers for ACT Work Keys test content and fairness

	Number of Reviewers	
Assessment Title	Content	Fairness
Reading for Information	13	8
Applied Mathematics	9	8
Applied Technology	14	15

Finally, ACT employed the process of matching items to the performance domain by virtue of a comparison by subject matter experts of job skill requirements versus Work Keys skill scales. The subject matter experts, usually individuals who were doing or had recently done the job profiled, were asked to classify job skill requirements in the following manner: "(a) not applicable to the job or fell below the Work Keys skill scale, (b) applicable but above the Work Keys skill scale, or (c) within the applicable Work Keys skill scale (i.e., content valid for the job)" (ACT, 1996, p. 11). The conclusion of the ACT staff after receiving the results of over 400 profiled jobs each for most skill areas was that:

> For all of the jobs and skills profiled, the vast majority of the jobs had skill requirements within the Work Keys skill scales. This strongly suggests that the Work Keys skill scales are content valid for large numbers of jobs. (ACT, 1996, p. 11)

Regarding criterion-related validation, Crocker and Algina (1986) state:

> The general design of a criterion-related validation study has the following steps:
> 1. Identify a suitable criterion behavior and a method for measuring it.
> 2. Identify an appropriate sample of examinees representative of those for whom the test will ultimately be used.
> 3. Administer the test and keep a record of each examinee's score.
> 4. When the criterion data are available, obtain a measure of performance on the criterion for each examinee.
> 5. Determine the strength of the relationship between test scores and criterion performance. (p. 224)

The Validity Supplement (ACT, 1996) contained criterion-related validation analysis results for only two of the three Work Keys assessment tests used in this investigation, Reading for Information and Applied Technology. No results were available in this supplement regarding Applied Mathematics. The study described in the Supplement (p. 12-13) in which the Reading for Information assessment was evaluated involved 47 workers across two companies and two types of jobs. The criterion behavior against which the Reading for Information assessment was compared was a newly developed general worker rating scale. ACT (1996, p. 13) concluded that "for the most part the study demonstrates good validity for Work Keys tests ... ." The average correlation and correlation range were 0.31 and 0.11 to 0.58 respectively (p. 13).

A second study reported in the Validity Supplement (1996, pp. 13-14) regarding the Applied Technology assessment yielded a correlation of 0.09 between the overall job performance ratings of 60 incumbent customer service technicians and their Applied Technology assessment test scores. Since ACT reported that no workers were rated as poor on their overall job performance, this analysis was considered to be "ineffective and inappropriate" for reasons of restriction of range. An alternate method using a 2 x 2 expectancy table, where workers were coded by job success and Work Keys test score relative to the level profiled for the job, is presented in Table 3.2. It appears from the results that the test was an inadequate indicator of success on the job.

Regarding construct validation, Crocker and Algina (1986) state that:

> Although construct validation evidence is typically assembled through a series of studies, the process generally contains the following steps:
> 1. Formulate one or more hypotheses about how those who differ on the construct are expected to differ on demographic characteristics, performance criteria, or measures of other constructs whose relationship to performance criteria has already been validated. These hypotheses should be based on an explicitly stated theory that underlies the construct and provides its syntactic definition.
> 2. Select (or develop) a measurement instrument which consists of items representing behaviors that are specific, concrete manifestations of the construct.
> 3. Gather empirical data which will permit the hypothesized relationships to be tested.
> 4. Determine if the data are consistent with the hypotheses and consider the extent to which the observed findings could be explained by rival theories or alternative explanations (and eliminate those if possible). (pp. 230-231)

Table 3.2. Percentage of examinees scoring at Level 6 on the Applied Technology test and rated as to success on the job

	Below Level	At or Above Level
Unsuccessful	0%	0%
Successful	88%	12%

ACT results in the Validity Supplement concerning the three Work Keys assessments used in this investigation were limited to internal consistency reliability and intercorrelational analyses. They (ACT, 1996, p. 15) reported internal consistency reliabilities for the "current" operational forms of the Work Keys assessment for level and total scores using Item Response Theory (IRT) based number correct scoring. The results for the three Work Keys assessments used in this investigation can be found in Table 3.3 (adapted from ACT Table 7).

Table 3.3. Reliability scores for Work Keys tests

	Reading for Information	Applied Mathematics	Applied Technology
Number of Items	30	30	32
Level Score Reliability	0.72	0.80	0.72
Total Score Reliability	0.79	0.85	0.77

ACT (1996) also reports the correlations among Work Keys skill areas (see Table 3.4). These correlations are deemed important because they "illustrate the degree to which the same or different constructs are being assessed" (p. 18). When discussing an intercorrelation of 0.56, ACT (1996) states that: "This level of correlation is low enough to indicate that somewhat separate skills are being measured, but high enough to indicate that these skills overlap" (p. 18).

The final group of intercorrelation statistics reported by ACT (1996) relevant to this investigation dealt with the comparison of the Work Keys Reading for Information, Applied Mathematics, and Applied Technology assessments with a company's existing array of selection tests. Applicable components of the company's tests included a mechanical reasoning test, a numerical reasoning test, and a teamwork/leadership test. To obtain data for the analysis, both sets of tests were administered to 100 college students in an introductory psychology class.

Table 3.4. Work Keys test intercorrelations--above the diagonal--and sample sizes--below the diagonal. This table is an adaptation of Table 9, page 18 of the 1996 ACT document Work Keys: Validity Supplement.

Work Keys Test	Reading for Information	Applied Mathematics	Applied Technology
Reading for Information	-----	0.56	0.41
Applied Mathematics	18,276	-----	0.48
Applied Technology	8,920	14,200	-----

Table 3.5 is an adaptation of Table 11, page 22 of the 1996 ACT document Work Keys: Validity Supplement. ACT (1996) interprets the above values as an indication of a "strong relationship between the Work Keys assessments and the reasoning tests, both verbal and numeric, and a moderate relationship with the teamwork/leadership test" (p. 21).

Table 3.5. Work Keys test intercorrelation statistics

	Mechanical Reasoning	Numerical Reasoning	Teamwork / Leadership
Reading for Information	0.50	0.49	0.33
Applied Mathematics	0.59	0.68	0.17
Applied Technology	0.57	0.52	0.29

Procedures

Work Keys instructions for this study are in Appendix A. Note: Specific Work Keys™ materials in this document are used with the permission of ACT.

Data Collection

The high schools chosen for the study were selected from an original list of approximately sixty schools. This original list was compiled based on the recommendations of Regional Tech Prep Coordinators and others knowledgeable about applied academics implementation efforts in the state. Initially it was thought that representative high schools could be chosen from each of the fifteen tech prep "regions" in the state and that stratified random samples of students enrolled in each of the two types of courses, applied and traditional, could be obtained from each of these schools. Stratified samples were felt to be desirable given the effort to minimize variation between the two groups of students resulting from gender, geography, and differing levels of academic ability or motivation. Upon closer examination, however, few of the sixty schools met all the key criteria necessary for inclusion in the study. For example, not all schools offered both an applied class and a comparable traditional class. Among schools offering both applied and comparable traditional classes, not all of them maintained separation between the teaching methods and materials (that is, some applied courses were "infused" or composites of applied and traditional methods and materials). In order to avoid problems associated with infused courses, an attempt was made to isolate and compare the extremes, 100% applied versus 100% traditional. This process allows one to highlight differences between the two curricula. In addition to the above, class size; availability of funds to purchase test materials; willingness of teachers, administrators, and students to participate in the study; and time schedules of participants, researchers, and the funding agency all played a role in the selection criteria. Based on the background and time frame surrounding the study, fewer schools and classes than originally intended were available for the study. The selection process resulted in the previously identified sample of schools.

Once schools were selected, the project team worked with each individual school to identify and schedule specific classes to take the Work Keys assessment tests. Researchers from Iowa State University (ISU) explained the project to staff at the school or district central administration and supplied all basic elements of information necessary for informed consent, either verbally or by written correspondence. Staff in the schools or school districts were responsible for subsequently obtaining informed consent from the parents of students who were to undergo testing. In one particular case a member of the ISU project staff was asked to submit a letter describing the project to the school district central administration; this letter would then be attached to a cover letter supplied by the school district and sent to the parents of the students targeted for Work Keys assessment tests. Participation in the study was strictly voluntary, but there seems to have been a high level of interest on the part of both the schools and the students since the majority of students in all classes targeted for testing did take the tests. A commitment was made to the schools, however, that only aggregate results from the investigation would be reported. In order to maintain the confidentiality of the participants, the results were compiled in such a manner as to prevent identification of individual school districts, individual schools, or individual students from the data.

The following data for all students in the target classes were collected during the 1995-1996 school year:

- High school
- Course type (applied or traditional)
- Course (math, communications, physics, etc.)
- Class
- Student

- Grade level (9 through 12)
- Student Cumulative High School Grade Point Average (0 to 4.00)
- Iowa percentile rank of the student's Iowa Test of Educational Development composite score (0 to 100)
- One or more of the three Work Keys test scores (the choice of Work Keys test(s) administered to each student depended upon the course in which the student was enrolled at the time of the study).

The Work Keys tests were administered during April or May at the end of the 1995-1996 school year.

Data Analysis

It is useful at this point to recall the original research questions since they play such a vital role in the development of the data analysis.

1. Is there a difference in student academic achievement for students who have completed applied academic courses in contrast to those who have completed comparable traditional academic courses?
2. Are the employability skills of students improved by their completion of applied academic courses in contrast to the employability skills of students who have completed comparable traditional academic courses?

The first order of business was to complete a descriptive analysis of the raw data. During this phase the two treatment groups were reviewed for obvious differences with respect to concomitant variables.

Following the descriptive analysis, exploratory data analysis was conducted to determine what form of analyses were best suited to the data. The first research question was then addressed using a paired sample test, either Student's t-test or Wilcoxon signed-rank method. In the case of the second research question, the initial effort focused on a model without covariates, again using a paired sample test. The analysis paired ***school x course Work Keys test means*** of students in applied and traditional courses. This section was included to address concerns previously expressed regarding the use of ANCOVA in a quasi-experimental designs. The reader later has the opportunity to contrast the results of tests on this model with subsequent hierarchical model results.

Development of the hierarchical model required one to choose among options in at least three areas: the first being the number of levels in the analysis; the second being the choice of parameters to be included in the model; and finally, the choice between fixed, random, and nonrandomly varying parameters.

Development of the Hierarchical Model

Development of the final form of the model followed a series of steps.

1. Univariate and bivariate relationships of variables that were considered for inclusion in the model were examined. As Bryk and Raudenbush (1992) mention: "Examination of the shape and scale of each variable provides a check on the quality of the data, identifies outlying observations, and may

suggest the need for a variable transformation" (p. 197). Correlational analysis was included in this section.

2. A fully unconditional three-level model was examined to clarify how variation in the Work Keys test results was partitioned across the different levels. The model structure to be examined consisted of students (Level 1) nested within classrooms (Level 2) nested within schools (Level 3). If the proportion of variance allocated to one level was small relative to the others, the model structure was simplified to a two-level model.
3. A multilevel analysis was performed. During this stage the questions of which predictor variables to include in the model and how their coefficients should be specified (fixed, random, nonrandomly varying) were answered.

Bryk and Raudenbush (1992) provided some guidance in building the models discussed in the third step above. They stated that:

> A natural temptation is to estimate a "saturated" Level-1 model--that is, where all potential predictors are included with random slopes--and then work backward deleting nonsignificant effects from the model. ...
>
> In general, we have found it more productive to use a "step-up" strategy. Assuming some external theoretical guidance that has defined a relatively small set of Level-1 predictors, we build up from univariate to bivariate to trivariate (and so on) models based on promising submodels. Often the best subset of Level-1 predictors can be identified through preliminary modeling using OLS Level-1 analyses. (p. 201)

Iversen (1991) provided some insight into the question of fixed versus random models. He mentioned that:

> There is not enough evidence in the literature to suggest that it is always necessary to use models with random regression coefficients. If there are strong, individual, group, and interaction effects present, there is reason

to believe that such effects will be found whether we use a model with random coefficients or one with fixed coefficients. It may be that using random coefficients is a better way to discover small effects, but small effects are not as substantively interesting.

Any result from a statistical analysis is due to two sources. One source is the data themselves, and the other source is the model used to analyze the data. ... We want the influences of the model we use to be as small as possible, which argues in favor of using models with fixed parameters when possible. (p. 77)

The Hierarchical Linear Model

Below is an example of a possible two-level conditional model. A more complete description of this model and its alternatives is provided in Appendix B.

General Level-1 Model: Within each classroom, one can model student employability skills (that is, Work Keys assessment test scores) as a function of the student-level predictors; here ITED composite score, gender, and grade level, plus a random student-level error:

$$Y_{(ij)} = \pi_{0(j)} + \pi_{1(j)}a_{1(ij)} + \pi_{2(j)}a_{2(ij)} + \pi_{3(j)}a_{3(ij)} + e_{(ij)}, \text{ where}$$

$Y_{(ij)}$ is the Work Keys test score of student i in classroom j.

$\pi_{0(j)}$ is the intercept for classroom j.

$\pi_{1(j)}a_{1(ij)}$ is the ITED term for student i in classroom j.

$\pi_{2(j)}\ a_{2(ij)}$ is the gender term for student i in classroom j.

$\pi_{3(j)}\ a_{3(ij)}$ is the grade level term for student i in classroom j.

$e_{(ij)}$ is a Level-1 random effect.

General Level-2 Model: Each of the regression coefficients in the above Level-1 model (including the intercept) can be viewed as either fixed, random, or nonrandomly varying. In addition each Level-1 coefficient may be predicted or modeled by some classroom-level characteristics.

An example of a fixed Level-1 coefficient is:

$$\pi_{0(j)} = \beta_{00}$$

An example of a random Level-1 coefficient is:

$$\pi_{1(j)} = \beta_{10} + r_{1(j)}$$

An example of a nonrandomly varying Level-1 coefficient is:

$$\pi_{2(j)} = \beta_{20} + \beta_{21}X_{1(j)}, \text{ where}$$

β_{00}, β_{10}, and β_{20} are all intercept terms.

$\beta_{21}X_{1(j)}$ is the term associated with curricula type (applied or traditional) used in classroom *j*.

$r_{1(j)}$ is a Level-2 random effect.

The techniques described in Bryk and Raudenbush (1992) were used to develop the final model. Regarding methods to guide model building at Level-1, they suggested that:

> In terms of a hierarchical analysis, two questions need to be addressed: (a) Is the fixed effect of [each Level-1 predictor] significant? and (b) Is there any evidence of slope heterogeneity ... [Level-1 coefficient variances greater than 0]? Statistical evidence of slope heterogeneity includes the point estimates, $\hat{\tau}_{qq}$, and the corresponding homogeneity test statistics (χ^2

> and likelihood-ratio tests ...). Also useful in this regard are the estimated reliabilities of the OLS intercepts and slopes.
>
> When the reliabilities become small (e.g., < 0.05), the variances we wish to estimate are likely to be close to zero ... . Inspection of the reliabilities may suggest that a random Level-1 coefficient be respecified as either fixed or nonrandomly varying. (p. 202)

Regarding methods to guide model building at Level-2, Bryk and Raudenbush (1992) suggested that:

> The most direct evidence of whether a Level-2 predictor should be included is the magnitude of its estimated effect and the related *t* ratio. Predictors with *t* ratios near or less than 1 are obvious candidates for exclusion from the model. (p. 212)

Methodological Assumptions

1. Conditional on a student's Level-1 predictor variables (such as grade level and gender), the within-class errors are normal and independent with class means of zero and equal variances across classes.
2. Whatever student-level predictors of employability skills (Work Keys assessment test results) are excluded from the model and thereby end up in the Level-1 error term, $e_{(ij)}$, are independent of a student's included Level-1 predictor variables.
3. The vector of residual class effects ($r_{0(j)}$, $r_{1(j)}$, $r_{2(j)}$, $r_{3(j)}$) are assumed multivariate normal, with mean vector (0, 0, 0, 0), variance-covariance matrix

$$\begin{bmatrix} \tau_{00} & \tau_{10} & \tau_{20} & \tau_{30} \\ \tau_{01} & \tau_{11} & \tau_{21} & \tau_{31} \\ \tau_{02} & \tau_{12} & \tau_{22} & \tau_{32} \\ \tau_{03} & \tau_{13} & \tau_{23} & \tau_{33} \end{bmatrix}.$$

4. Whatever class-level predictors of the intercept and student-level coefficients are excluded from the model and thereby end up in the Level-2 error terms, for example $r_{0(j)}$, are independent of the included class-level (Level-2) predictor variables, such as curricula type.
5. The error at Level-1 is independent of the Level-2 error terms.

Methodological Limitations

One must be careful regarding the number of Level-1 coefficients specified as random in the model. Bryk and Raudenbush (1992) state:

> One cannot be definitive about how many random effects can be specified because the maximum will depend on several factors: the magnitude of the variance components, the degree of intercorrelation among the random effects, and other characteristics of the data. (p. 203)

They do go on to provide one example based on a subsample of data from the High School and Beyond survey. They mention that:

> For example, using the High School and Beyond data with about 60 students per school and 160 schools, we have found that three random coefficients plus a random intercept is about as rich a model as the data can sustain. (p. 203)

A limitation that was previously mentioned in Chapter 2 concerns the failure to specify certain types of covariates. Such omissions can lead to serious biases in Level-2 predictors. For example, if a student's socioeconomic status is related to employability skills, and if students in applied and traditional curricula differ significantly on socioeconomic status, and if socioeconomic status is ignored in the model, then the estimates of the effect of curricula on employability skills will be biased. There are at least two ways to remove bias once a decision has been made to include the covariate; however one can never be sure that all relevant covariates have been included. In addition, data samples are rarely large enough to include investigation of all possible covariates.

Previously mentioned errors of measurement (in Chapter 2) can also bias estimates of the coefficients. According to Bryk and Raudenbush (1992): "The degree of bias depends on the explanatory power of the true predictor, the degree of unreliability of its measurement, and the intercorrelations among predictors" (p. 217).

Nonnormality of Level-1 errors is not a concern when estimating Level-2 coefficients, according to Bryk and Raudenbush (1992), but "will introduce bias into standard errors at both levels and therefore into the computation of confidence intervals and hypothesis tests. Little is currently known about the direction and severity of such effects" (p. 210).

One other issue concerns the validity of inferences when sample sizes are small. Bryk and Raudenbush (1992) addressed this issue as well in their book.

> In unbalanced cases we rely on large sample theory. Estimates of the fixed effects and their standard errors depend upon point estimates of each of the variance and covariance components in the model. Because of the mutual dependence of the point estimates of the fixed effects and the point estimates of the variance-covariance components, the exact sampling distributions of the resulting estimators are unknown. However, when maximum likelihood estimation is used, the large-sample properties of the maximum likelihood estimators are known. The question in this section is: How well does the large sample distribution theory work?

> For any particular application, the answer depends upon the inferences sought. (p. 220)

For reasons of time and space, the interested reader is referred to Bryk and Raudenbush (1992, pp. 220-229) for a discussion of large sample distribution theory and its potential impact on estimation of fixed effects, random effects, and variance-covariance components.

Summary

The investigation used a sample of Iowa high school students, grades 9 through 12, enrolled in specific applied or traditional curricula. The goal of Work Keys assessment tests, to provide a measure of employability skills, is admirable; however, additional validity and reliability studies would be beneficial. Since gain scores are particularly susceptible to measurement errors, it seemed to be inadvisable to use them as a measure of curricula effectiveness here. A hierarchical linear model was the model of choice. The investigation included both unconditional (no predictor variables) and conditional models. Although there are well-documented concerns as to the use of control variables in quasi-experimental studies, an attempt was made to include such variables at certain points in the analysis. These attempts were made, hopefully, in keeping with the advice offered by Pedhazur (1982) when he stated that: "the conduct of such research, indeed all scientific research, requires sound theoretical thinking, constant vigilance, and a through understanding of the potential and limitations of the methods being used" (p. 525).

CHAPTER 4. RESULTS

Overview

Findings are reported in this chapter. The chapter begins with an examination of the sample demographics. Information at all three levels (student, class, and school) that would allow identification of specific students, classes, or schools was purposely withheld for reasons of confidentiality. The study looked at the percentage of total variance divided among the levels, but information that would allow one to "rank" performance was not included.

Following a description of the sample, characteristics of the variables were investigated. This included examination of the univariate distributions and bivariate relationships of variables during exploratory data analysis (EDA). Multiple graphical procedures were employed to review the data prior to the paired sample analyses and the hierarchical modeling. Residual analysis was also employed during the hierarchical modeling process. These sections were needed to provide insight into the tenability of model assumptions; the validity of which can affect the legitimacy of inferential statistics developed from those models. For example, Student's t-tests may be used if the data come from a normal (or nearly normal) distribution without outliers. If the data contain outliers or do not appear to come from a normal distribution, then robust or nonparametric methods should be used.

The section titled statistical data analysis includes two components: the first covering questions regarding academic achievement of students in applied courses versus those in traditional courses; and the second, addressing questions concerning the impact of curricula type (applied versus traditional) on employability skills. Demographic differences were explored using one of several possible methods, such as Student's t-test or Wilcoxon signed-rank test, depending on the shape of the distribution and the presence of outliers; while the impact of curricula type was investigated using Hierarchical Linear Model (HLM) techniques as described in Bryk and Raudenbush (1992). Several software packages were used during the investigation; S-PLUS for Windows (Version 3.3, Release 1) by MathSoft, HLM for Windows (Version 4.01.01) by Scientific Software International, and SPSS for Windows (Release 6.1.3) by SPSS. Under "Testing the Hypotheses", one will find results for each hypothesis posed in the previous chapter. The final section is a summary of the chapter.

Description of the Sample

Missing data are almost always a concern in education research and this study was no exception. The original sample of 1,321 students resulted in full matrix data for 842 students after eliminating series with missing or obviously erroneous data points. A full matrix for this investigation included information regarding school, type of course (applied or traditional), course, class within course, student gender, student grade in school, student cumulative grade point average (GPA), Iowa Tests of Educational Development score (Iowa percentile rank), the name of the Work Keys test, and the Work Keys test score. The full matrix data were collected from students in 88 classes at nine Iowa high schools. The locations of the nine high schools are shown in Figure 4.1.

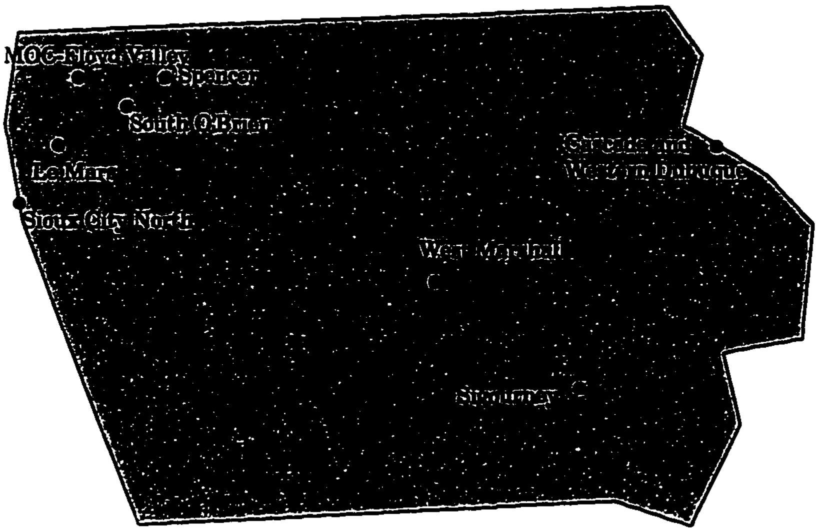


Figure 4.1. Sites of high schools included in sample

Since some students took more than one Work Keys test, a total of 1,265 full matrix data series were available for analysis. A summary of the number of students taking single and multiple tests is provided in Table 4.1. One should note, however, that full matrix data were not always required for purposes of this study, particularly when analyzing demographic characteristics and when generating Level 2 (Class) and Level 3 (School) data for inclusion in the HLM portion of the study. These data are identified as vector data series. Table 4.2 shows the number of students included in each vector data series. In all cases, vector data includes data regarding school, type of course, course, class, and the variable shown in Table 4.2.

Table 4.1. Number of students taking Work Keys tests (full matrix)

Work Keys test(s)	Number of students
Applied Math only	369
Applied Technology only	165
Reading for Information only	79
Applied Math & Applied Technology	18
Applied Math & Reading for Information	10
Applied Technology & Reading for Information	7
All three tests	194
Total	842

Table 4.2. Number of students included in vector data series

Variable	Number of students
Student gender	1,270
Student grade	1,270
Student GPA	1,255
ITED composite score	1,008
Applied Mathematics Work Keys test score	790
Applied Technology Work Keys test score	526
Reading for Information Work Keys test score	414

A sample of 591 students at eight different high schools took the Applied Math Work Keys test and provided all other requested data. The difference between the number (790) shown in Table 4.2 and the 591 sample size stated above is predominately a result of data series eliminated due to missing ITED data. The listings of the students who took the test, the number of classes among which these students were divided, and the mean and standard deviation of the number of students taking the test in each class, are shown below in Table 4.3. The courses most closely associated with the subject matter in the Applied Math Work Keys test are in bold italics.

Table 4.3. Class size statistics for students taking Applied Mathematics Work Keys tests (full matrix)

		Number of Students in Class taking Test		
Course	Number of Classes	Total	Class Mean	Class Standard Deviation
Applied Math I	***11***	***110***	***10.00***	***5.67***
Traditional Math I	***12***	***122***	***10.17***	***9.05***
Applied Math II	***8***	***73***	***9.13***	***4.12***
Traditional Math II	***10***	***75***	***7.50***	***4.99***
Applied Communications	5	48	9.60	5.46
Traditional English	8	72	9.00	3.38
Applied Technology I	9	26	2.89	1.62
Physics	1	8	8.00	-----
Applied Biol./Chem.	4	8	2.00	1.41
Traditional Biol./Chem.	4	49	12.25	4.50

A sample of 384 students at six different high schools took the Applied Technology Work Keys test and provided all other requested data. The listings of the students who took the test, the number of classes among which these students were divided, and the mean and standard deviation of the number of students taking the test in each class, are shown below in Table 4.4. The courses most closely associated with the subject matter in the Applied Technology Work Keys test are in bold italics.

Table 4.4. Class size statistics for students taking Applied Technology Work Keys tests (full matrix)

		Number of Students in Class taking Test		
Course	Number of Classes	Total	Class Mean	Class Standard Deviation
Applied Math I	4	6	1.50	0.58
Traditional Math I	0	0	-----	-----
Applied Math II	1	3	3.00	-----
Traditional Math II	0	0	-----	-----
Applied Communications	4	48	12.00	2.94
Traditional English	9	73	8.11	4.28
Applied Technology I	***11***	***86***	***7.82***	***2.23***
Physics	***9***	***110***	***12.22***	***3.80***
Applied Biol./Chem.	4	8	2.00	1.41
Traditional Biol./Chem.	4	50	12.50	4.80

A sample of 290 students at five different high schools took the Reading for Information Work Keys test and provided all other requested data. The listings of the students who took the test, the number of classes among which these students were divided, and the mean and standard deviation of the number of students taking the test in each class, are shown below in Table 4.5. The courses most closely associated with the subject matter in the Reading for Information Work Keys test are in bold italics.

Table 4.5. Class size statistics for students taking Reading for Information Work Keys tests (full matrix)

		Number of Students in Class taking Test		
Course	Number of Classes	Total	Class Mean	Class Standard Deviation
Applied Math I	4	6	1.50	1.00
Traditional Math I	1	1	1.00	-----
Applied Math II	3	16	5.33	2.52
Traditional Math II	1	1	1.00	-----
Applied Communications	***7***	***71***	***10.14***	***2.67***
Traditional English	***11***	***108***	***9.82***	***3.40***
Applied Technology I	6	17	2.83	2.23
Physics	2	11	5.50	3.54
Applied Biol./Chem.	4	8	2.00	1.41
Traditional Biol./Chem.	4	51	12.75	5.12

A series of figures and tables conclude the descriptive section of the chapter. The figures included in this section are examples taken from the full complement of over 70 descriptive histograms and boxplots in Appendix C.

Appendix C includes graphs covering grade, gender, GPA, ITED score, and Work Keys score variables compiled over three levels--student, class, and school. Each of these variables compares "applied" students versus "traditional" students in specific subgroups. The first series of Level 1 histograms in Appendix C include all students enrolled in applied courses versus all students enrolled in traditional courses; followed by histograms of students enrolled in specific applied courses versus their counterparts in specific traditional courses. The second series of Level 1 histograms in Appendix C include data on grade, gender, GPA, ITED score, and Work Keys score variables compiled only for those students who scored below the minimum skill level assessed on each of the three Work Keys tests included in the study. These graphs were generated to allow examination of student attributes for potential patterns in the important subgroup of students who did not meet minimum levels of "employability skills". The Level 2 graphs follow the same pattern as the first series of Level 1 graphs with the exception that class means are charted rather than individual student results. Boxplots are used to compare school-level data collected on applied versus traditional student groups; that is, Level 3 data are school means for all students enrolled in applied courses versus school means for all students enrolled in traditional courses.

Figures 4.2 through 4.6 allow the reader to make comparisons of the ITED distributions for students enrolled in applied versus traditional courses. In Figure 4.2, all students in the study enrolled in any of the applied courses are compared with all students in the study enrolled in any of the traditional courses. Figures 4.3 through 4.6 partition the above groups along class lines. Figures 4.9 through 4.13 allow class mean ITED comparisons; while Figure 4.14 includes ITED data compiled at the school level.

Tables 4.6 through 4.8 summarize the gender makeup of the sample. Table 4.6 shows that the split between male and female students in the sample, with the exception of the Applied Technology subgroup, is relatively even. Table 4.7 shows that when broken down along applied and traditional lines, however, males outnumber females in applied courses while the opposite is true in traditional courses. Table 4.8 ties specific preparatory course enrollment to the Work Keys test taken and examines the gender makeup. For example, of the 159 students enrolled in either Applied Math I or II who took the Applied Math Work Keys test, 80 are female and 79 are male. On the traditional side, one can use this table to determine that of the 110 students enrolled in Physics who took the Applied Technology test, 35 are female and 75 are male.

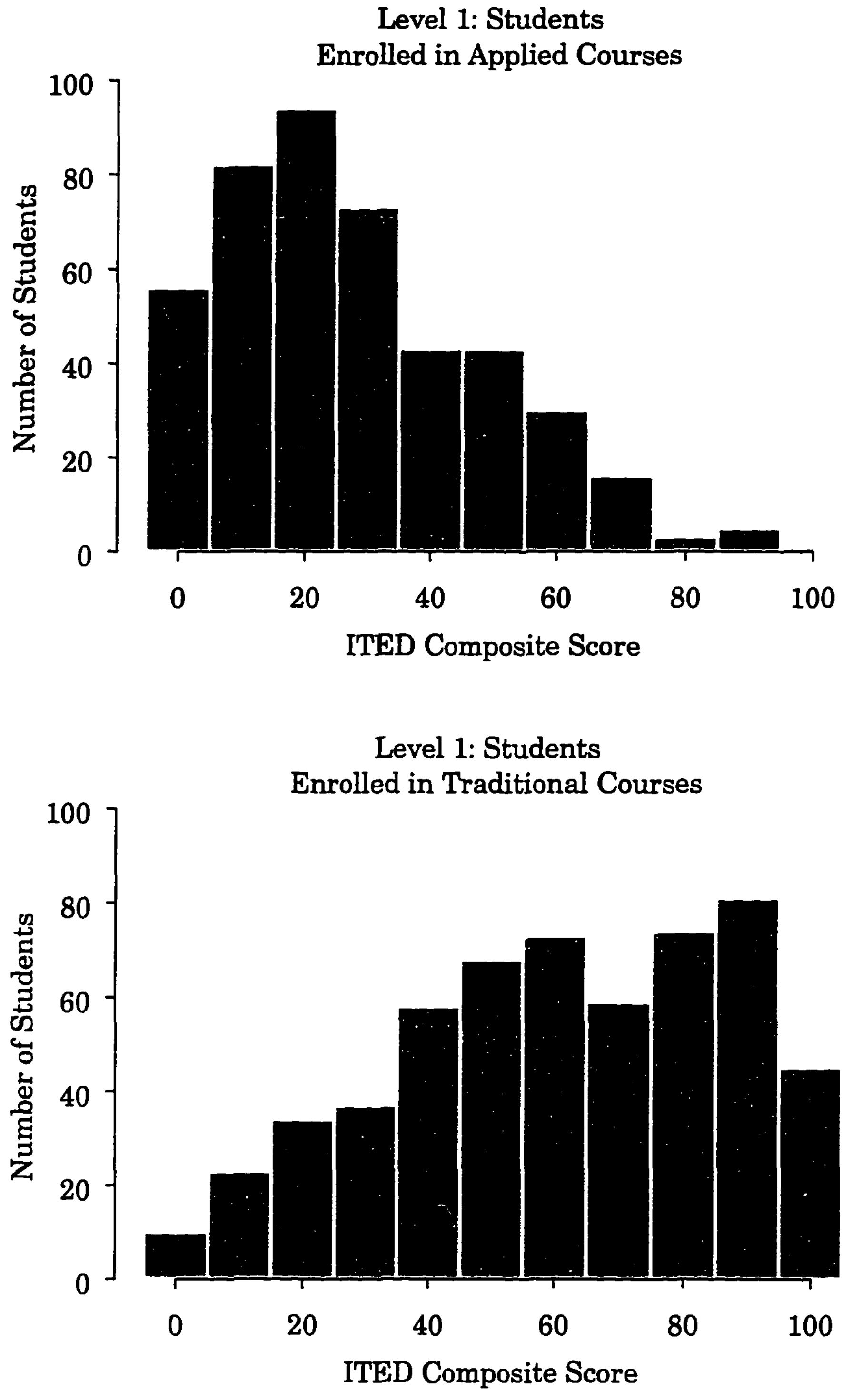
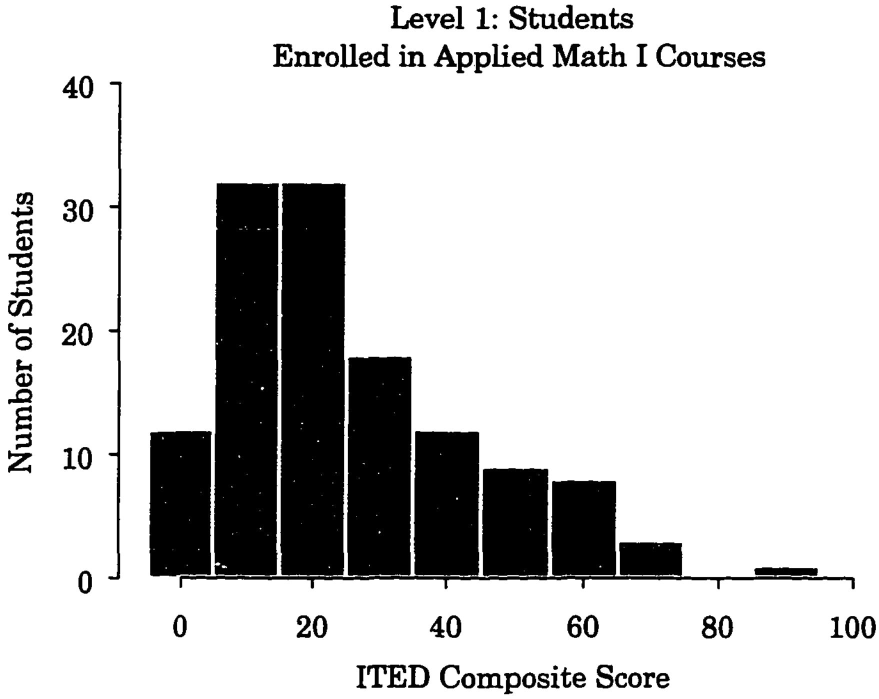


Figure 4.2. Histograms comparing "applied" versus "traditional" students' ITED score (vector data)



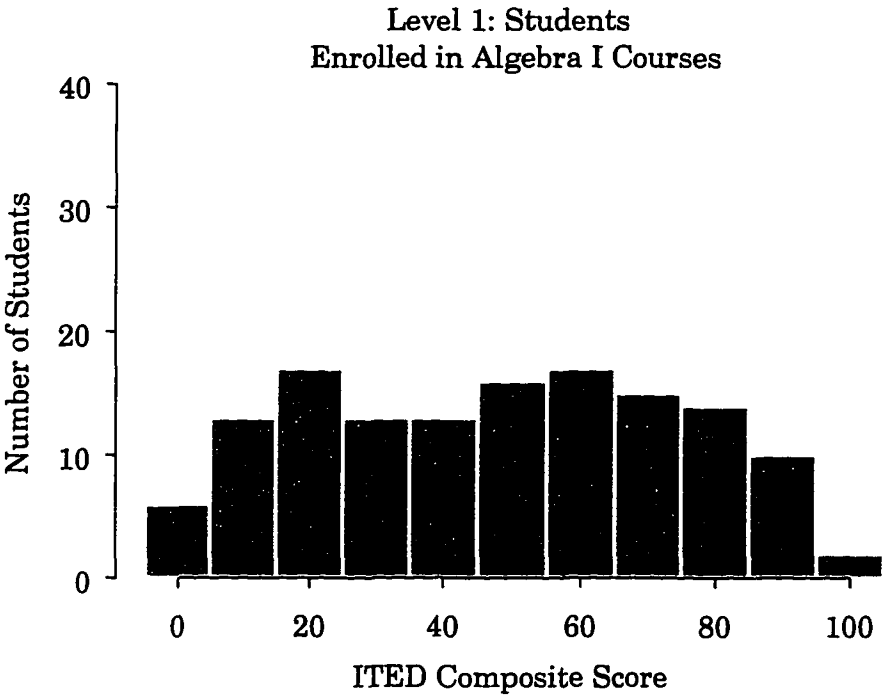
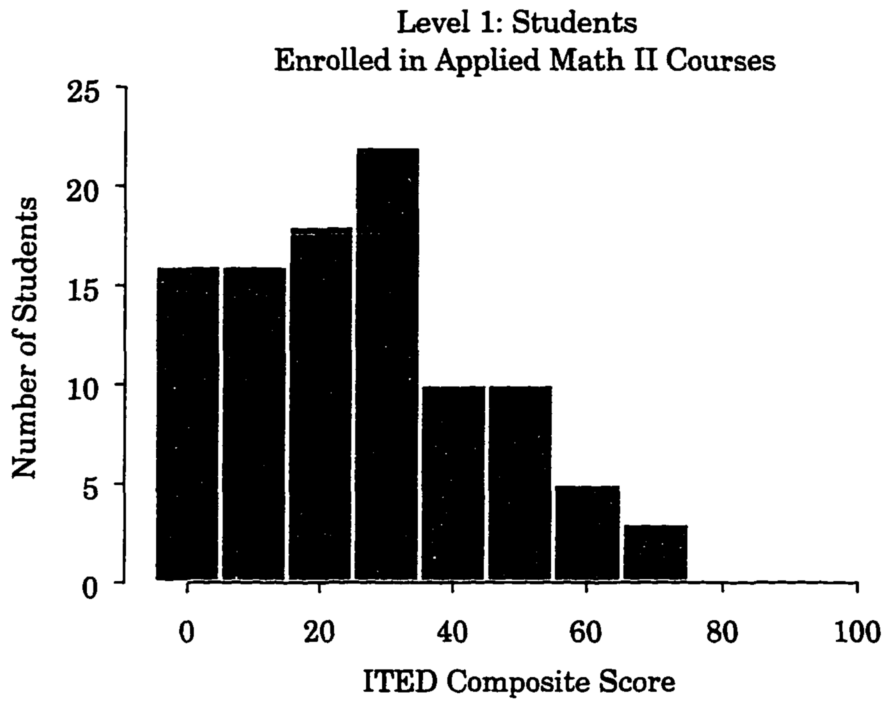


Figure 4.3. Histograms comparing Applied Math I versus Algebra I students' ITED score (vector data)



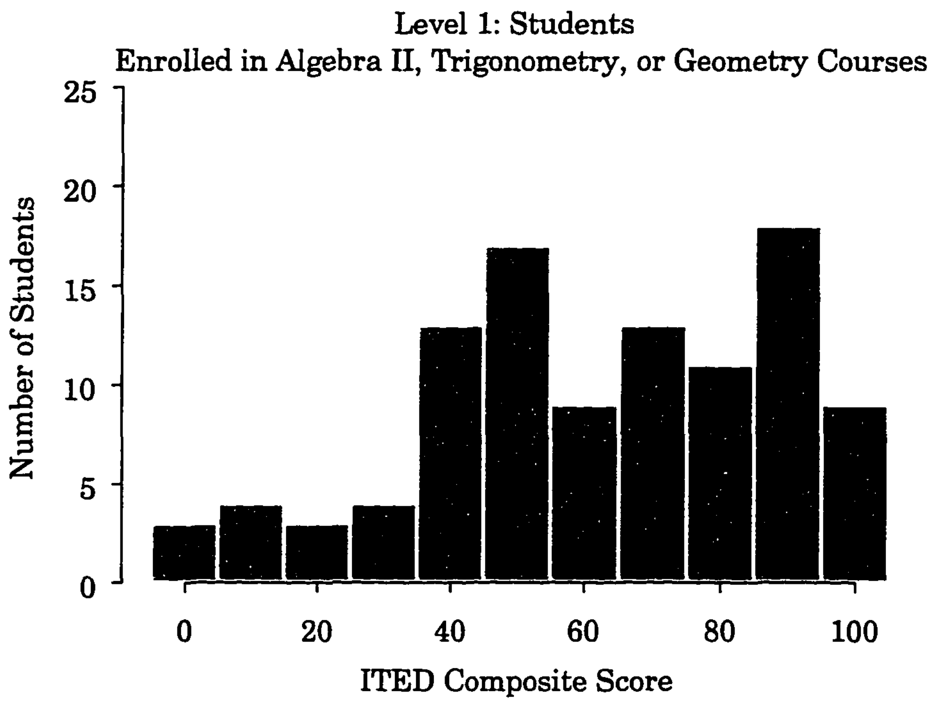


Figure 4.4. Histograms comparing Applied Math II versus Traditional Math II students' ITED score (vector data)

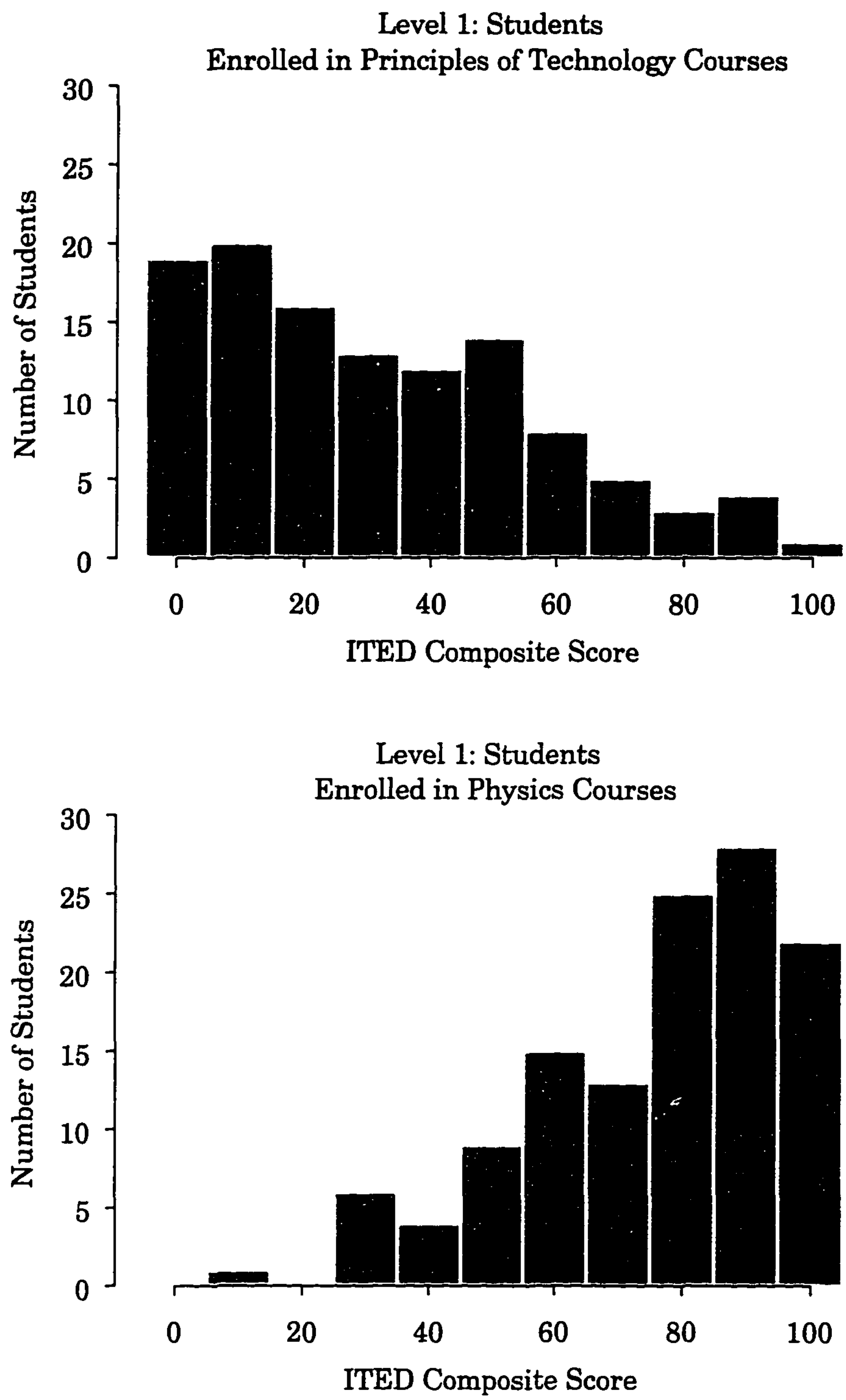


Figure 4.5. Histograms comparing Principles of Technology versus Physics students' ITED score (vector data)

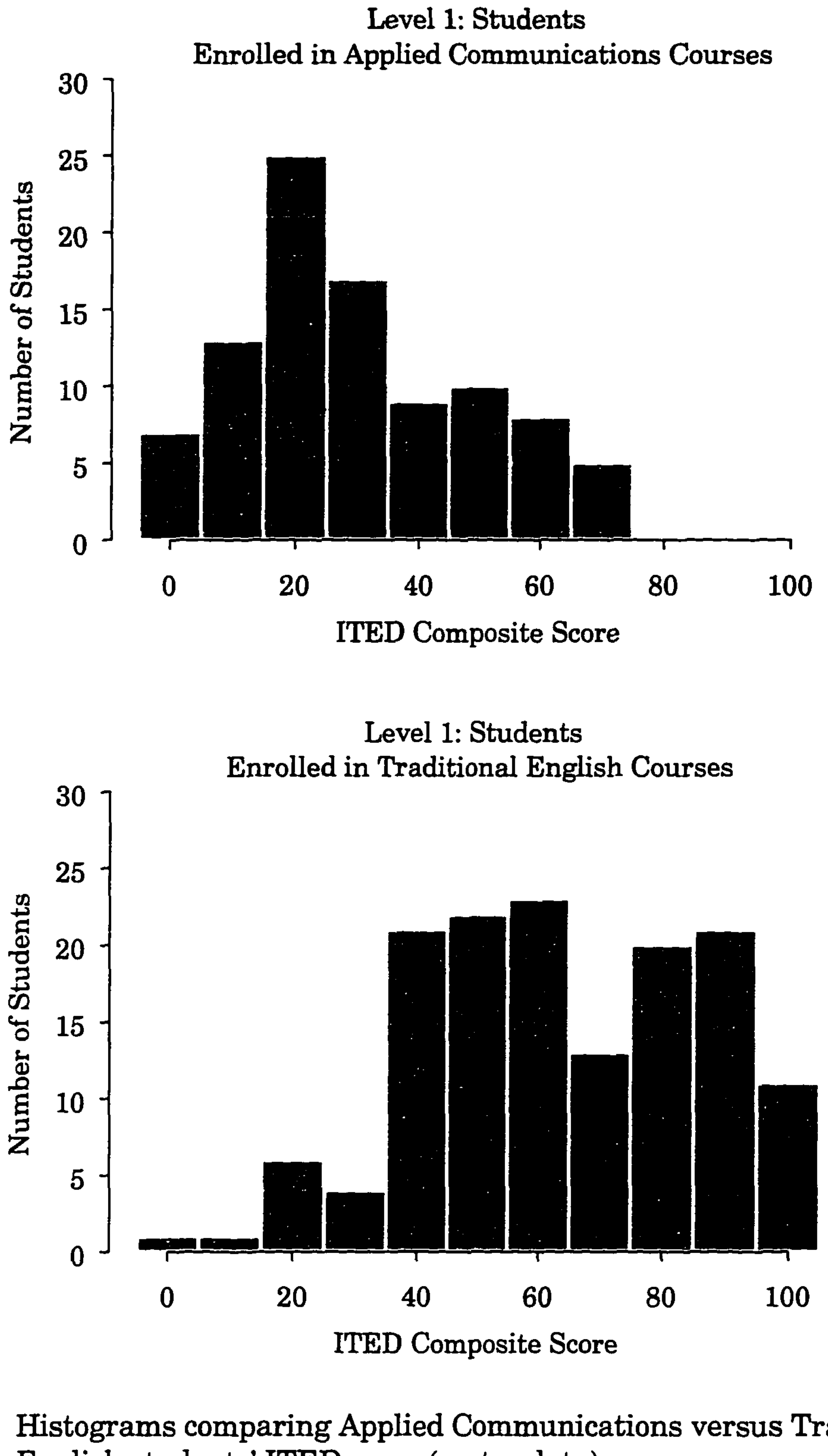


Figure 4.6. Histograms comparing Applied Communications versus Traditional English students' ITED score (vector data)

Table 4.6. Work Keys test versus Gender frequency table for all students (full matrix)

All Students	Gender		
Work Keys Test	Female	Male	Totals
Applied Math	295	296	591
Applied Technology	152	232	384
Reading for Information	148	142	290
Totals	595	670	1,265

Table 4.7. Work Keys test versus Gender frequency table for students enrolled in applied and traditional curricula (full matrix)

Applied Students	Gender		
Work Keys Test	Female	Male	Totals
Applied Math	105	160	265
Applied Technology	32	119	151
Reading for Information	35	83	118
Totals	172	362	534
Traditional Students	Gender		
Work Keys Test	Female	Male	Totals
Applied Math	190	136	326
Applied Technology	120	113	233
Reading for Information	113	59	172
Totals	423	308	731

Table 4.8. Work Keys test versus Gender frequency table for students with complete data enrolled in courses specific to the Work Keys test taken (e.g., only Physics students who took the Applied Technology test are included as Traditional Students in "Applied Technology")

Applied Students: Matched Courses & Test	Gender		
Work Keys Test	Female	Male	Totals
Applied Math	85	98	183
Applied Technology	12	74	86
Reading for Information	23	48	71
Totals	120	220	340

Traditional Students: Matched Courses & Test	Gender		
Work Keys Test	Female	Male	Totals
Applied Math	105	92	197
Applied Technology	35	75	110
Reading for Information	74	34	108
Totals	214	201	415

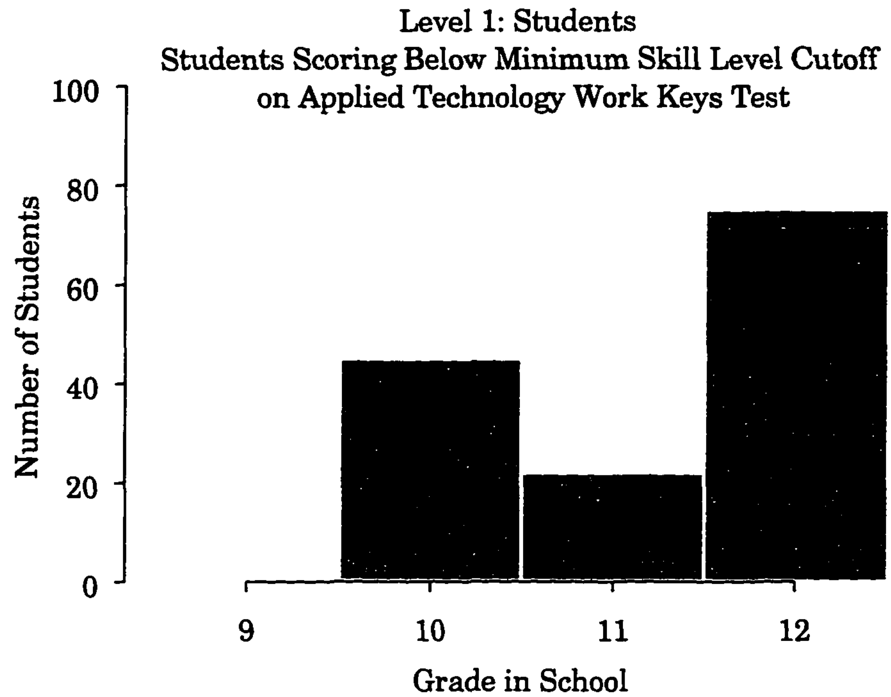
One other area of interest related to student, or Level 1, demographics is the split between students meeting or exceeding the minimum skill level cutoff score of 3 on the three Work Keys tests and those who did not. Table 4.9 presents a detailed summary of student results for all three tests. Figures 4.7 and 4.8 provide histograms of grade, gender, ITED score, and GPA for all students who scored below the minimum skill level score on the Applied Technology test.

Table 4.9. Frequency table by course enrollment of students scoring above and below minimum skill level cutoff on Work Keys test (vector data)

Applied Mathematics	Test Score		Student	*0's as %*
Course	0	≥ 3	Totals	*of Total*
Applied Math	8	208	216	*3.7%*
Traditional Math	2	239	241	*0.8%*
Applied Other	6	164	170	*3.5%*
Traditional Other	4	159	163	*2.5%*
Totals	20	770	790	*2.5%*

Applied Technology	Test Score		Student	*0's as %*
Course	0	≥ 3	Totals	*of Total*
Principles of Technology	43	54	97	*44.3%*
Physics	18	105	123	*14.6%*
Applied Other	81	69	150	*54.0%*
Traditional Other	76	80	156	*48.7%*
Totals	218	308	526	*41.4%*

Reading for Information	Test Score		Student	*0's as %*
Course	0	≥ 3	Totals	*of Total*
Applied Communications	6	64	70	*8.6%*
Traditional English	1	110	111	*0.9%*
Applied Other	16	118	134	*11.9%*
Traditional Other	7	92	99	*7.1%*
Totals	30	384	414	*7.2%*



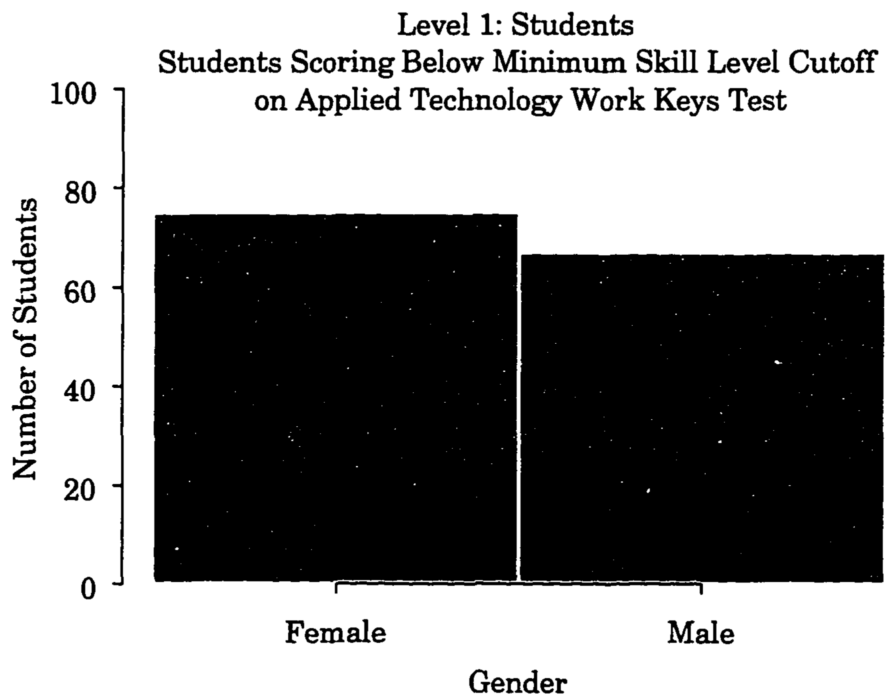
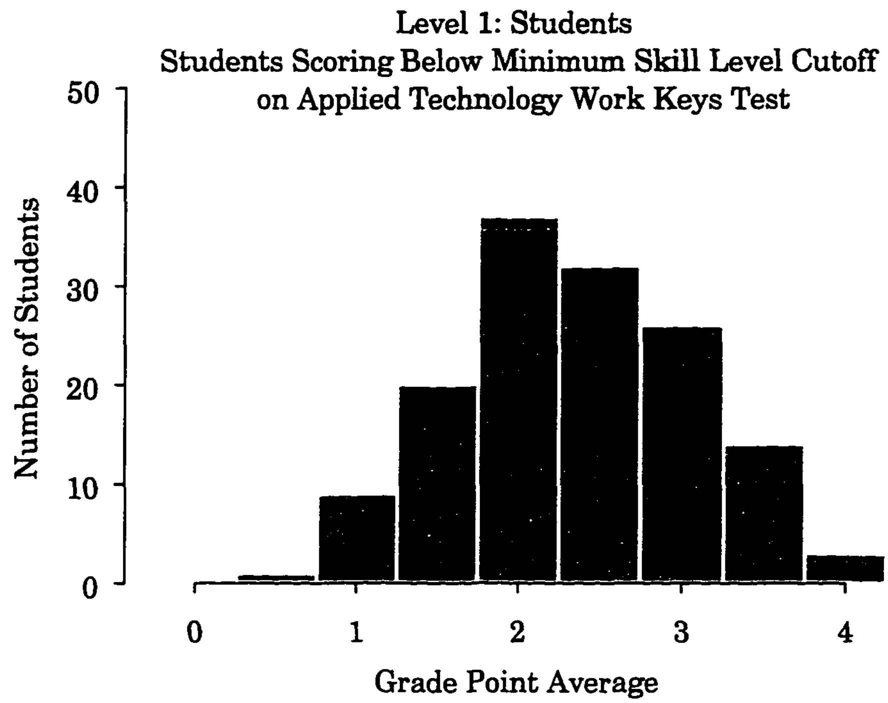


Figure 4.7. Grade and Gender histograms of students scoring below the minimum skill level of 3 on AT Work Keys test (full matrix data)



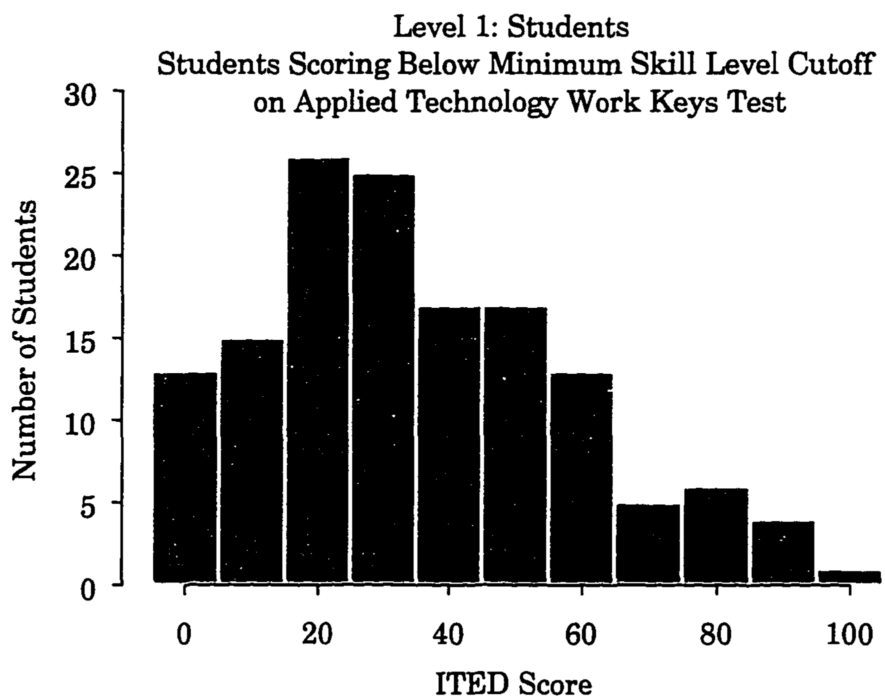
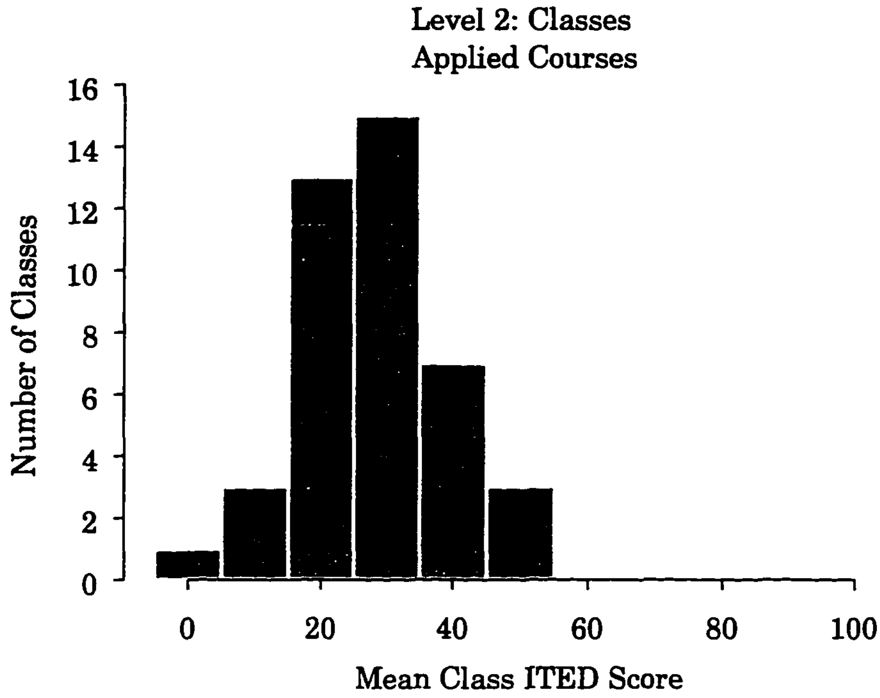


Figure 4.8. GPA and ITED histograms of students scoring below the minimum skill level of 3 on AT Work Keys test (full matrix data)



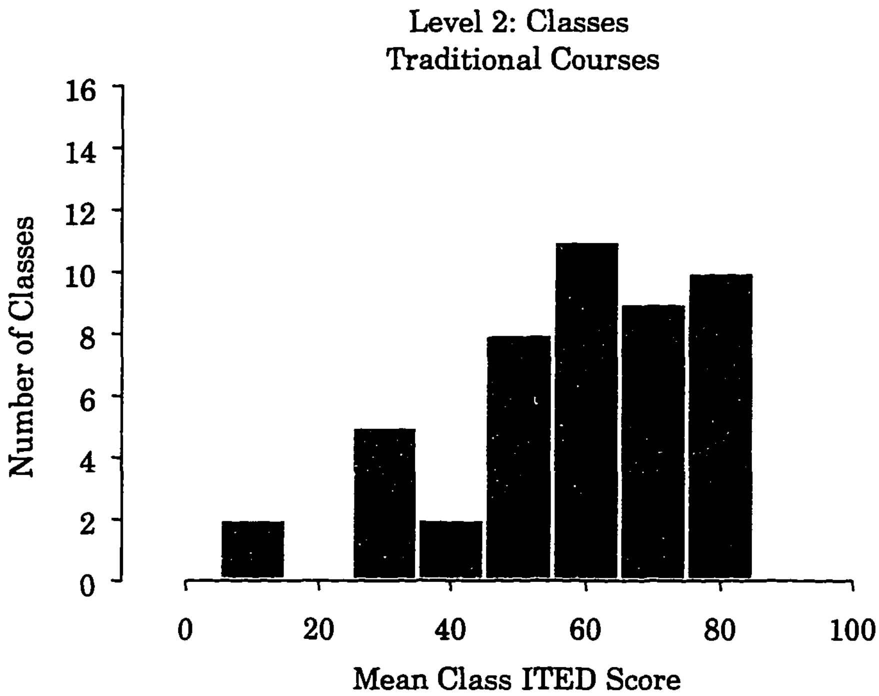
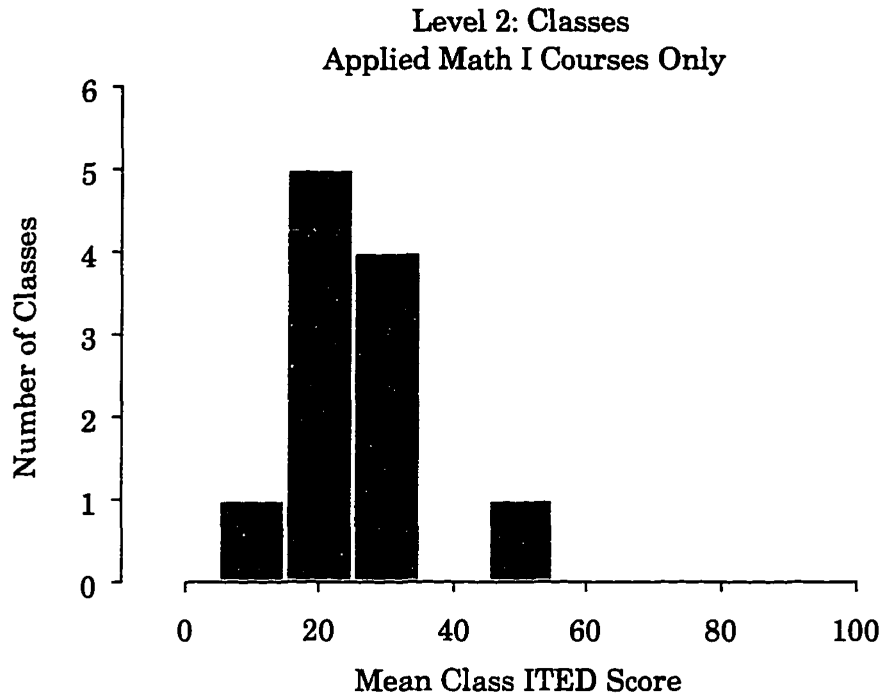


Figure 4.9. Histograms comparing mean class ITED score of applied classes versus traditional classes (vector data)



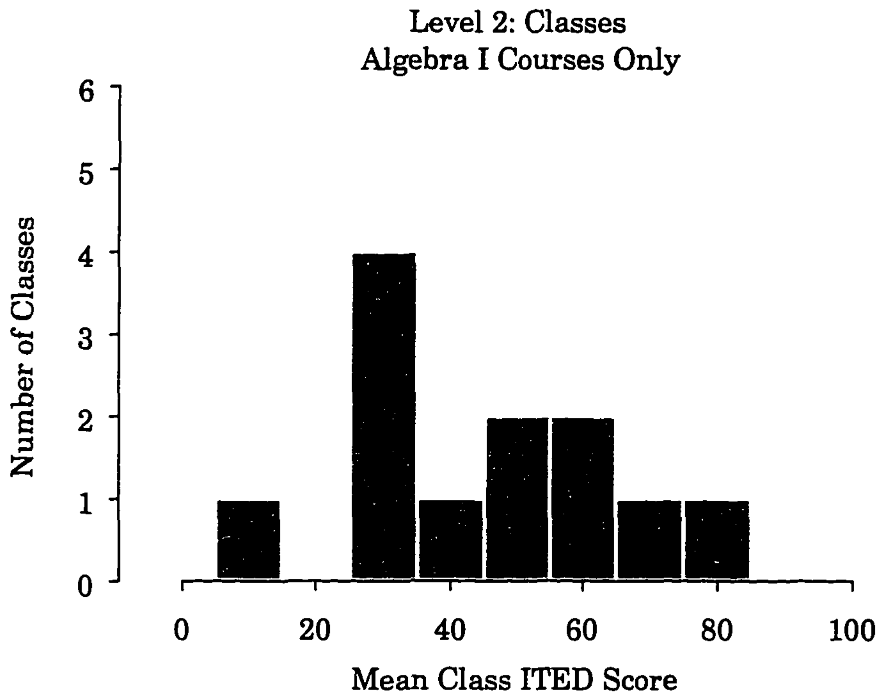


Figure 4.10. Histograms comparing mean class ITED score of Applied Math I versus Algebra I classes (vector data)

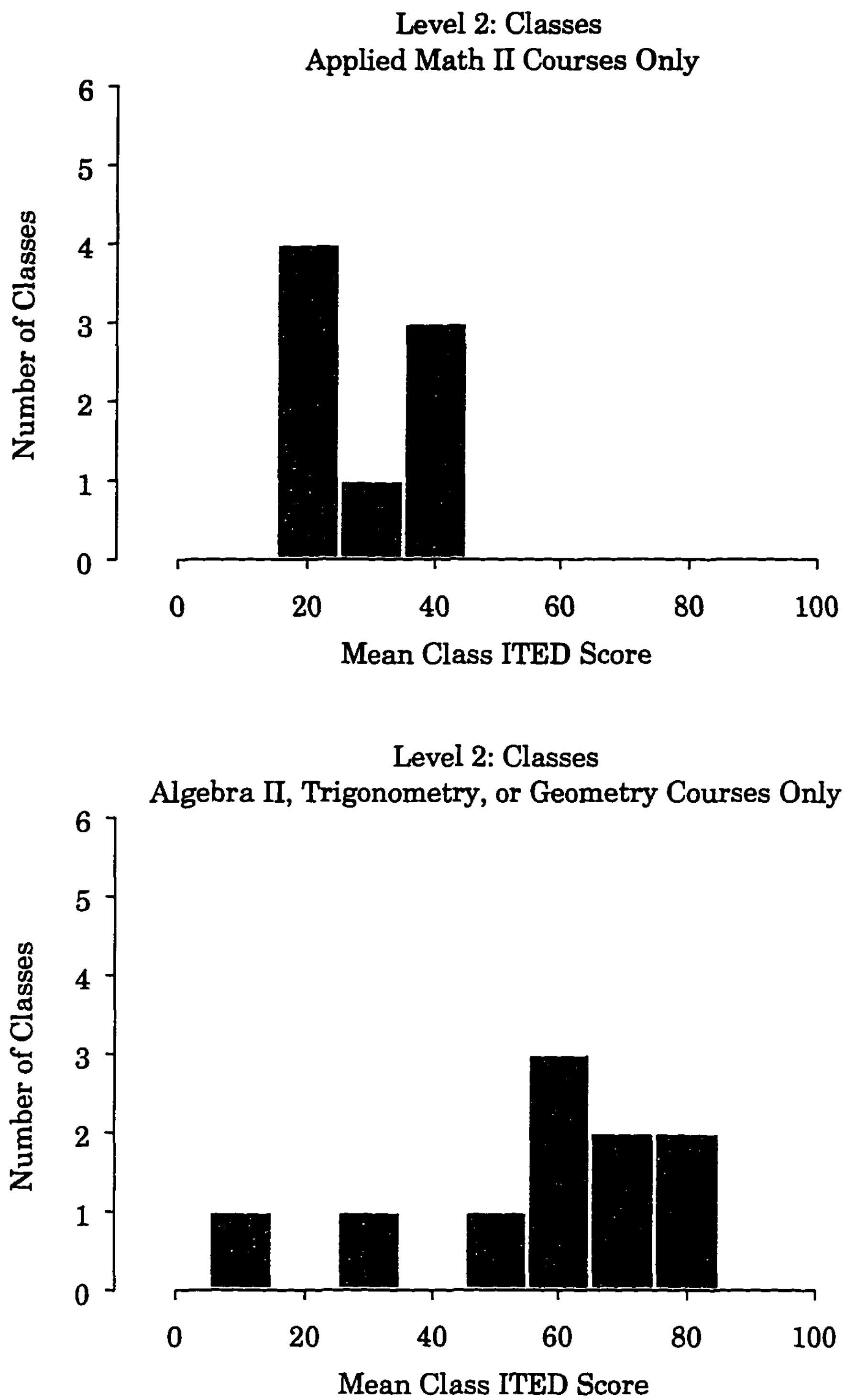
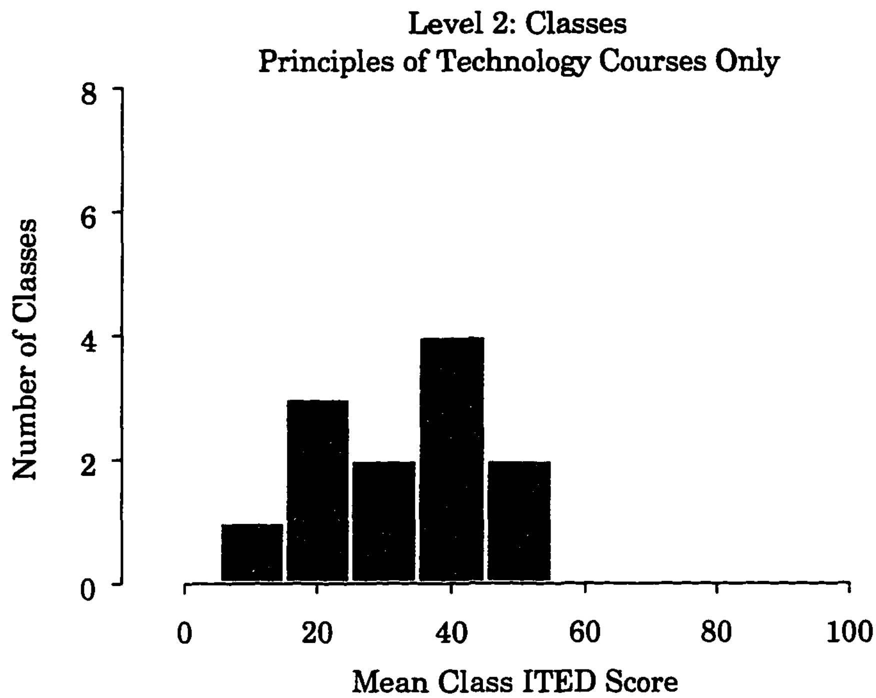


Figure 4.11. Histograms comparing mean class ITED score of Applied Math II versus Traditional Math II classes (vector data)



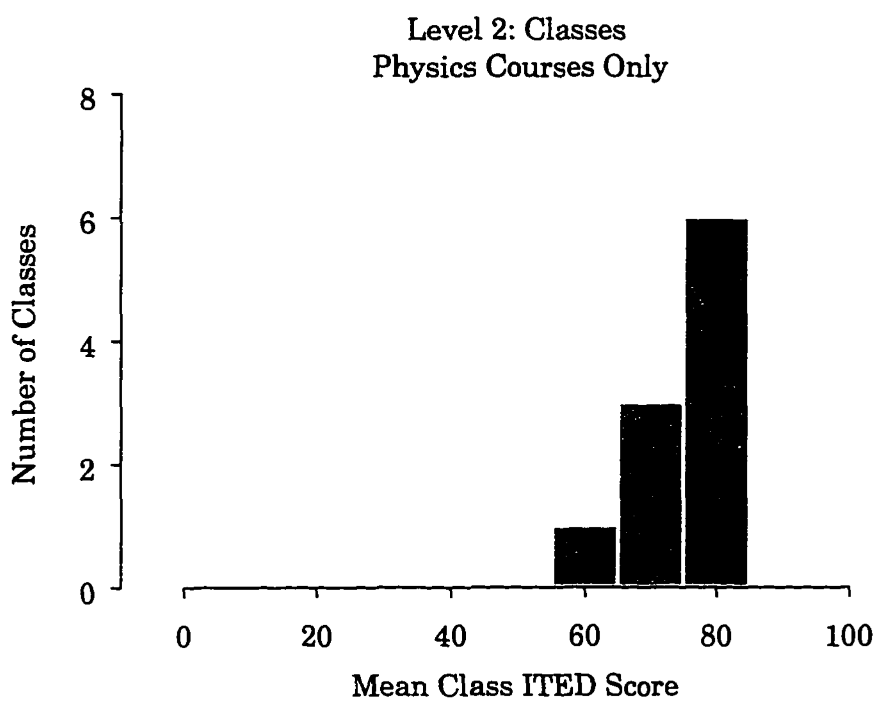


Figure 4.12. Histograms comparing mean class ITED score of Principles of Technology versus Physics classes (vector data)

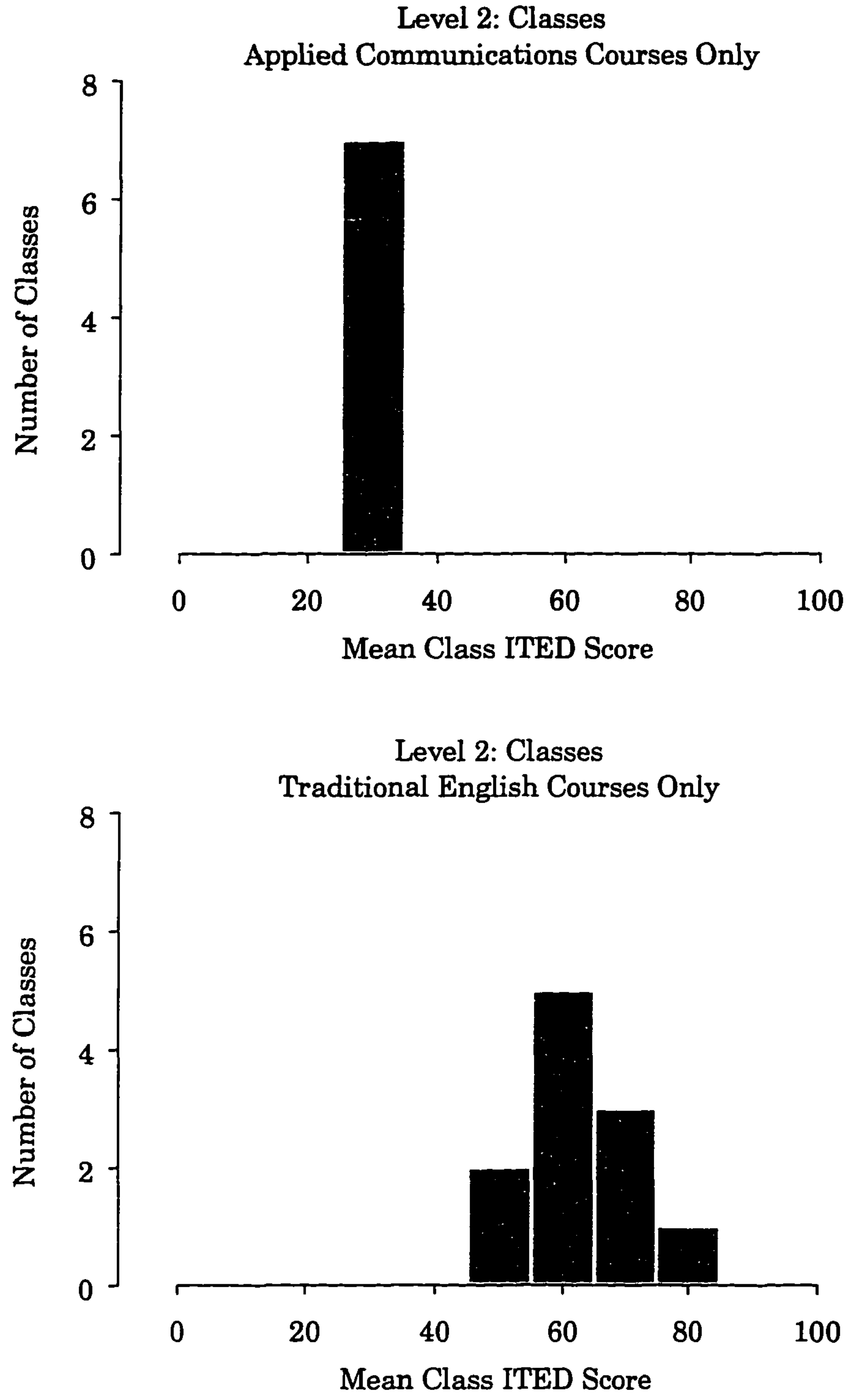


Figure 4.13. Histograms comparing mean class ITED score of Applied Communications versus English classes (vector data)

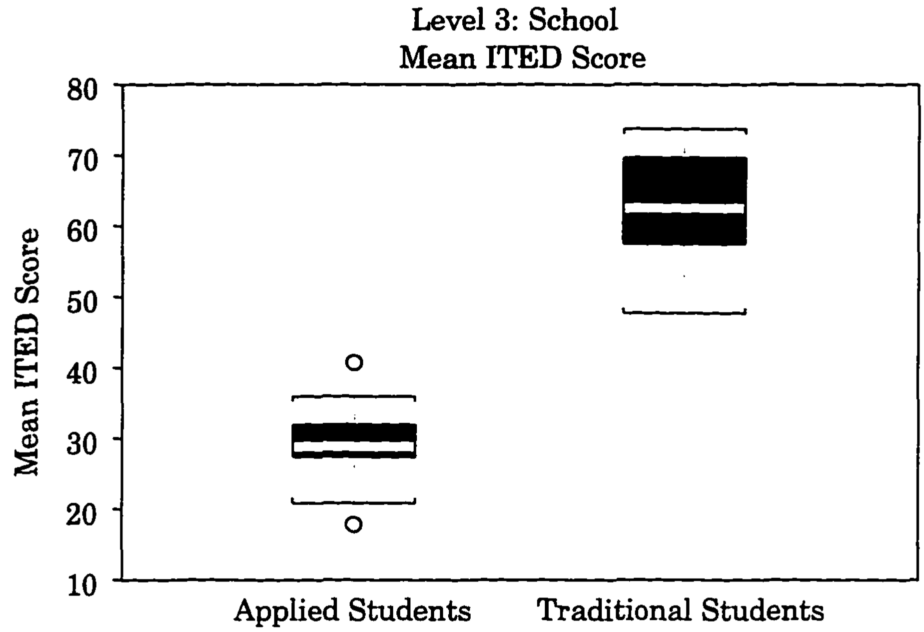


Figure 4.14. Boxplot comparing Level 3 mean ITED score for applied versus traditional students (vector data)

For those readers who may not be familiar with boxplots, the edges of the "box" are located at the first (25%) and third (75%) quartile values, thus the box covers the interquartile range (IQR) or middle 50% of the data. The line inside the box is located at the median data value. The ends of the "whiskers" are located at smallest and largest non-outliers, where non-outliers are defined as falling within a span 1.5 IQRs below the first interquartile value and 1.5 IQRs above the third interquartile value. Data values falling outside the whiskers are defined as potential outliers and are plotted individually as small ovals.

Figure 4.14 indicates all school mean ITED scores for traditional course students exceed the highest school mean ITED score for applied course students.

Exploratory Data Analysis

Student's t-test is a commonly used method for statistical inference. This test, however, relies on two critical assumptions:

1. The data have a common normal distribution with mean μ and variance σ^2.
2. The observations are independent.

If these assumptions do not hold, then one should use methods that are robust against deviations from the model. The Wilcoxon signed-rank method is one method robust against deviations from the assumption of normality.

According to Statistical Sciences (1995):

> You can get a pretty good picture of the shape of the distribution generating your data, and also detect the presence of outliers, by looking at the following collection of four plots: a *histogram*, a *boxplot*, *density plot*, and a *normal qqplot*.
>
> ... Density plots are essentially smooth versions of histograms, which provide smooth estimates of population frequency, or probability density curves, ...
>
> A normal qqplot (or quantile-quantile plot) consists of a plot of the ordered values of your data versus the corresponding quantiles of the standard normal distribution, i.e., a normal distribution with mean zero and variance one. If the qqplot is fairly linear, your data are reasonably Gaussian [normal]; otherwise they are not.
>
> Of these four plots, the histogram and the density plot give you the best picture of the distribution shape, while the boxplot and the normal qqplot give the clearest display of outliers. (Chapter 3, pp. 6-7)

Figures 4.15 through 4.18 are Exploratory Data Analysis (EDA) plots generated to determine whether the Student's t-test or the Wilcoxon signed-rank method should be used to answer the first (two-part) research question.

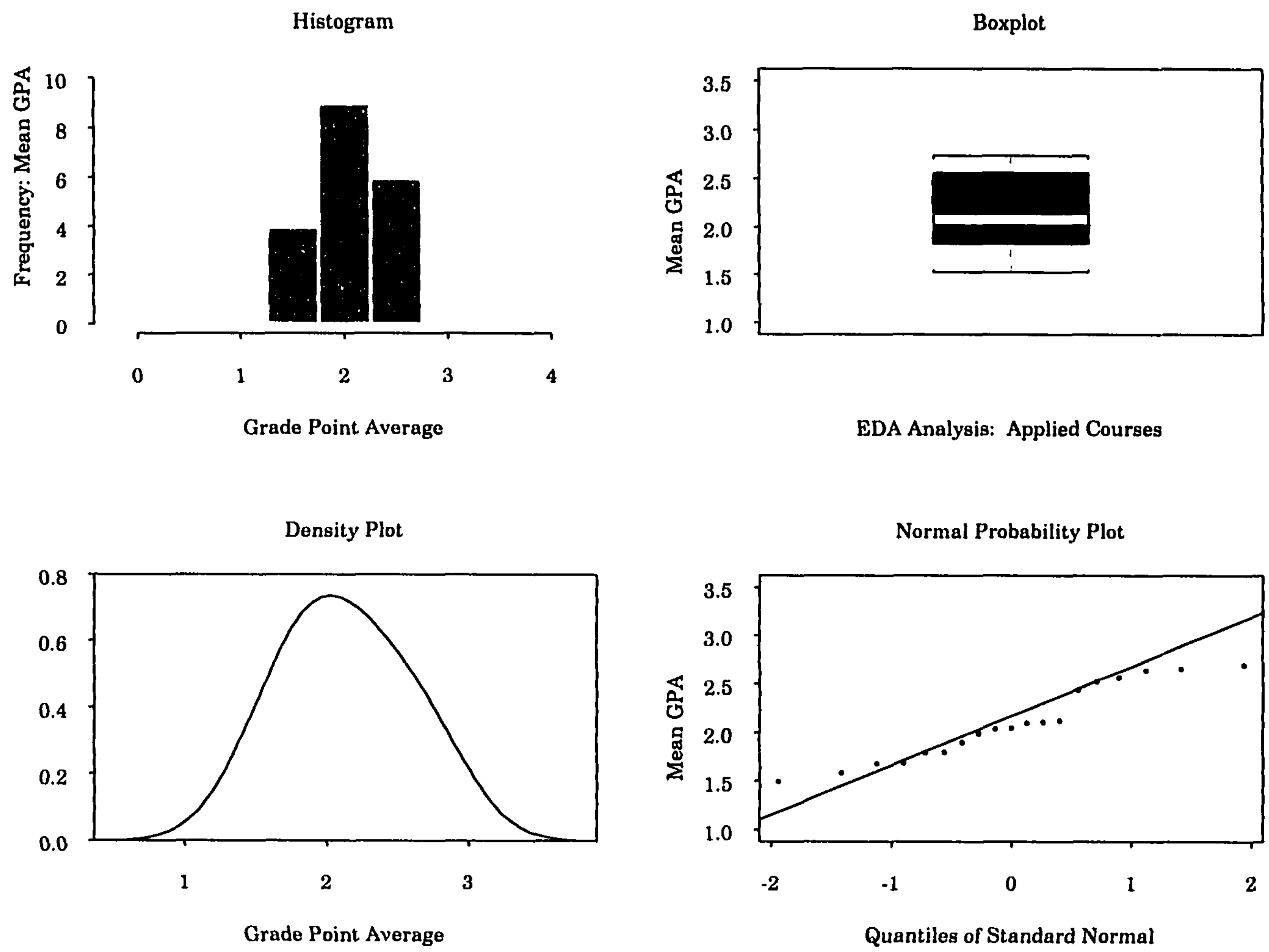


Figure 4.15. Exploratory Data Analysis plots of course x school mean GPA for students enrolled in applied courses

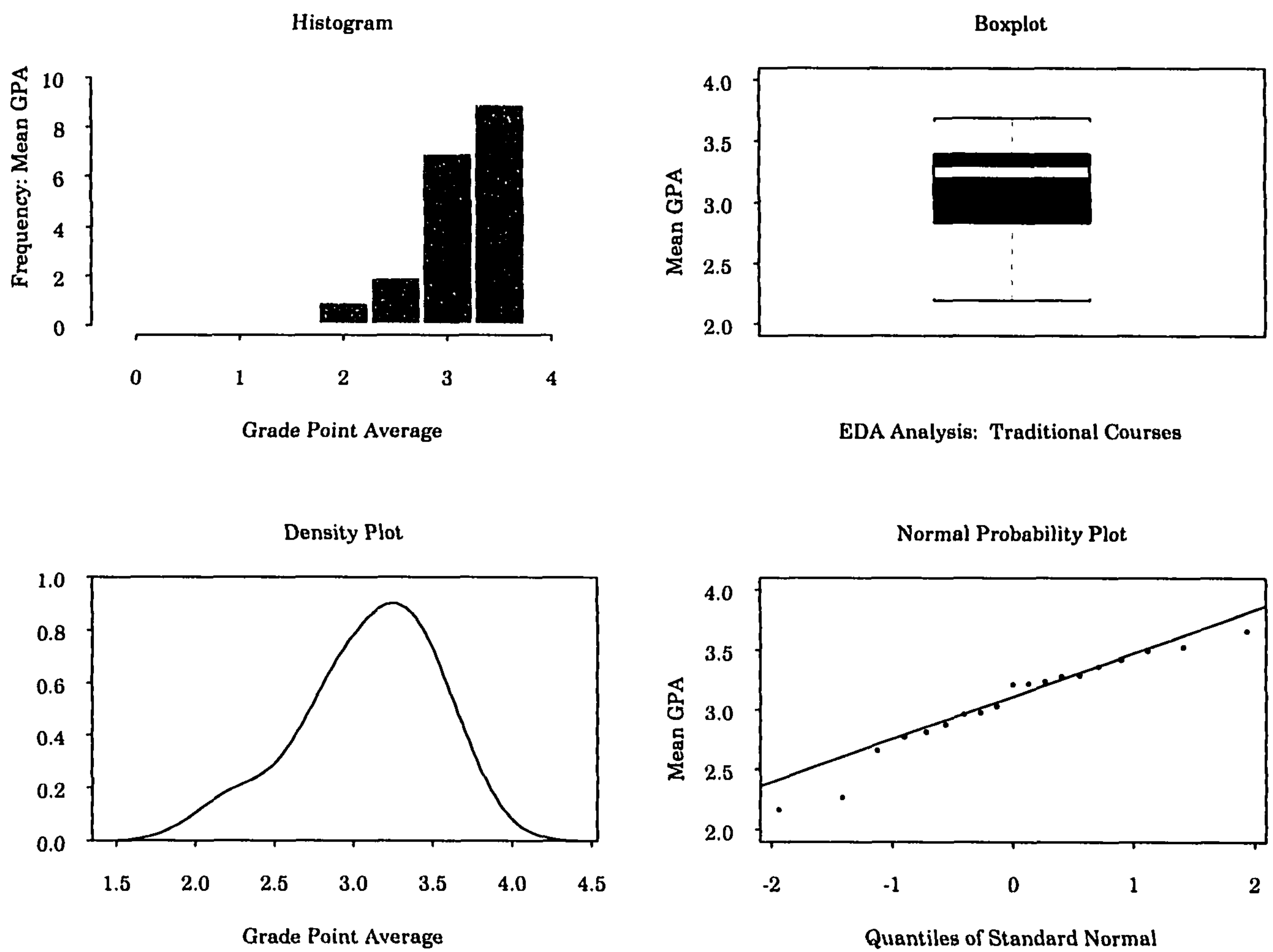


Figure 4.16. Exploratory Data Analysis plots of course x school mean GPA for students enrolled in traditional courses

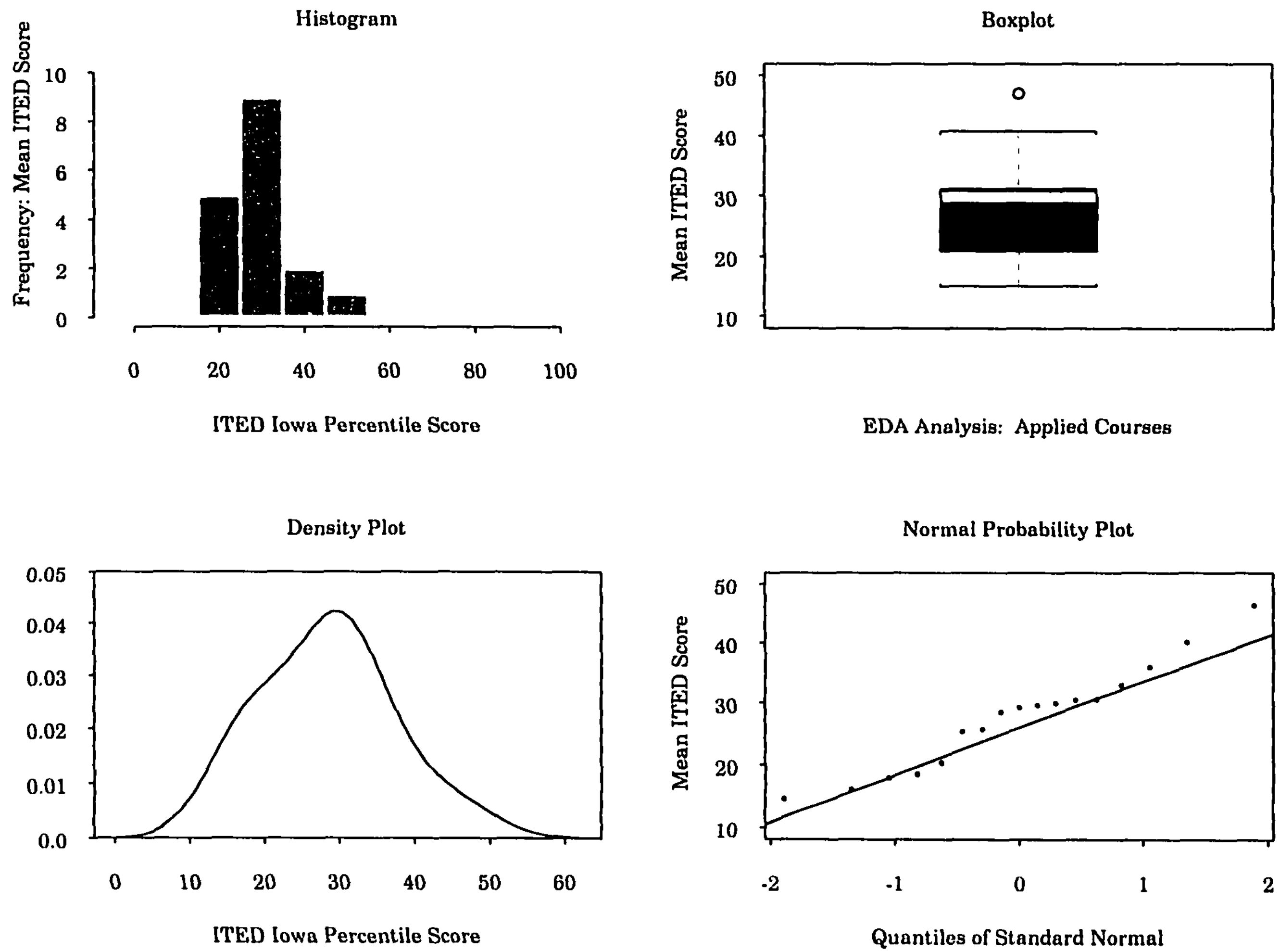


Figure 4.17. Exploratory Data Analysis plots of course x school mean ITED scores for students enrolled in applied courses

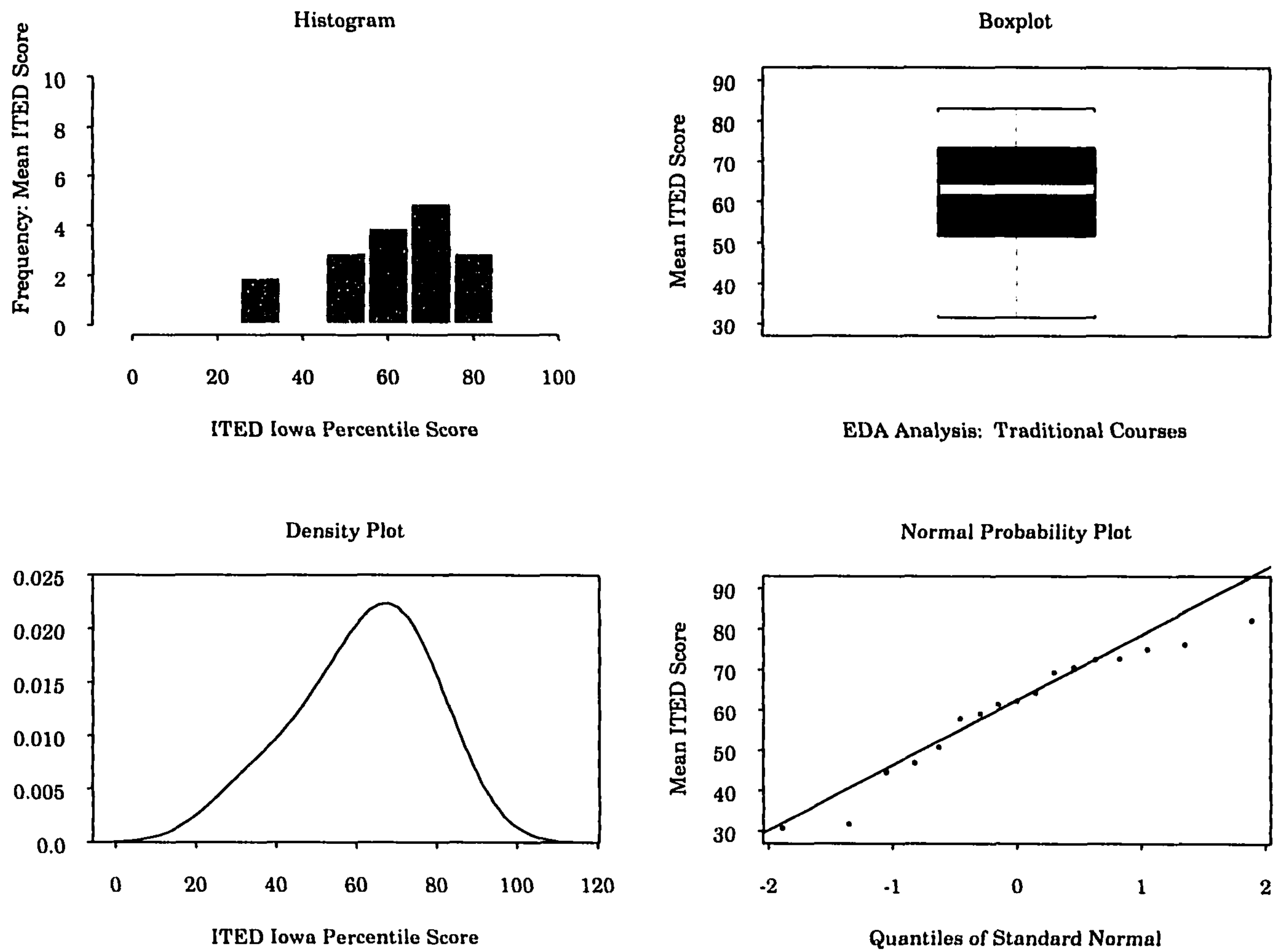


Figure 4.18. Exploratory Data Analysis plots of course x school mean ITED scores for students enrolled in traditional courses

Recall that the first question addresses academic achievement differences by evaluating differences in course mean grade point averages and course mean ITED scores. Although there are no obvious outliers visible in the GPA boxplots, there do appear to be some questionable points in the tails of the distribution as seen on the normal probability plots. In addition, the course mean GPA distributions, as seen in the density plots, are somewhat skewed and the fact that the distribution for traditional students is truncated on the high side is clearly visible in the histogram. The ITED distributions present similar concerns to those raised by the GPA density and normal probability plots. There is also an outlier present in the boxplot for students enrolled in applied courses. The EDA plots would seem to indicate that use of the Wilcoxon method is the more prudent course of action.

The next series of EDA plots , Figure 4.19 through Figure 4.24, cover course mean Work Keys test scores. Based on these plots, one could use the Student's t-test for the Applied Math data, but the Reading for Information data clearly requires the Wilcoxon method. The choice of analysis method for the Applied Technology data is less clear, however since the lower tail of the distribution for students enrolled in traditional courses shows some departure from normality on both the density and normal probability plots, the Wilcoxon method is recommended here as well.

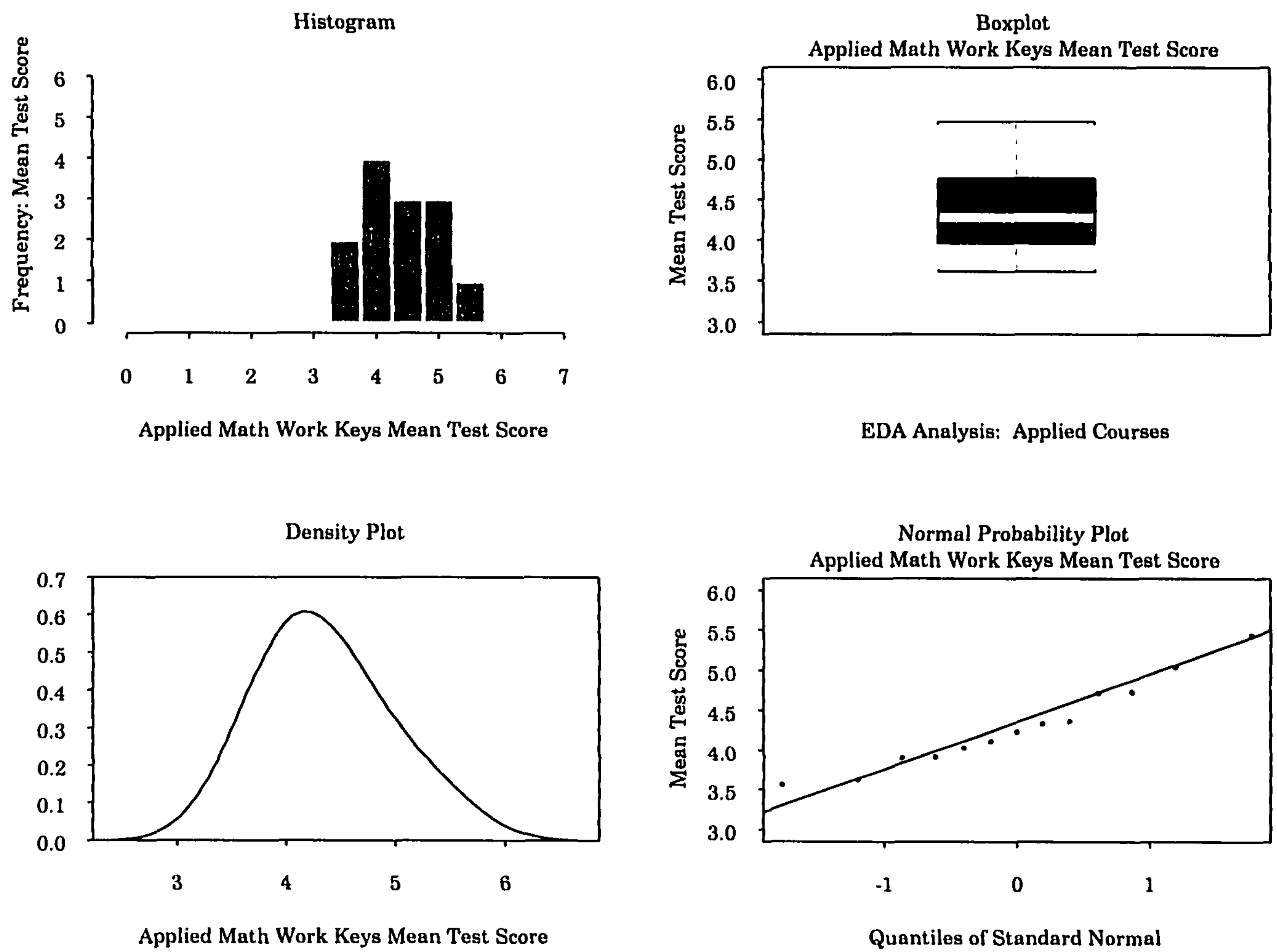


Figure 4.19. Exploratory Data Analysis plots of course x school mean Applied Math Work Keys test scores for students enrolled in applied courses

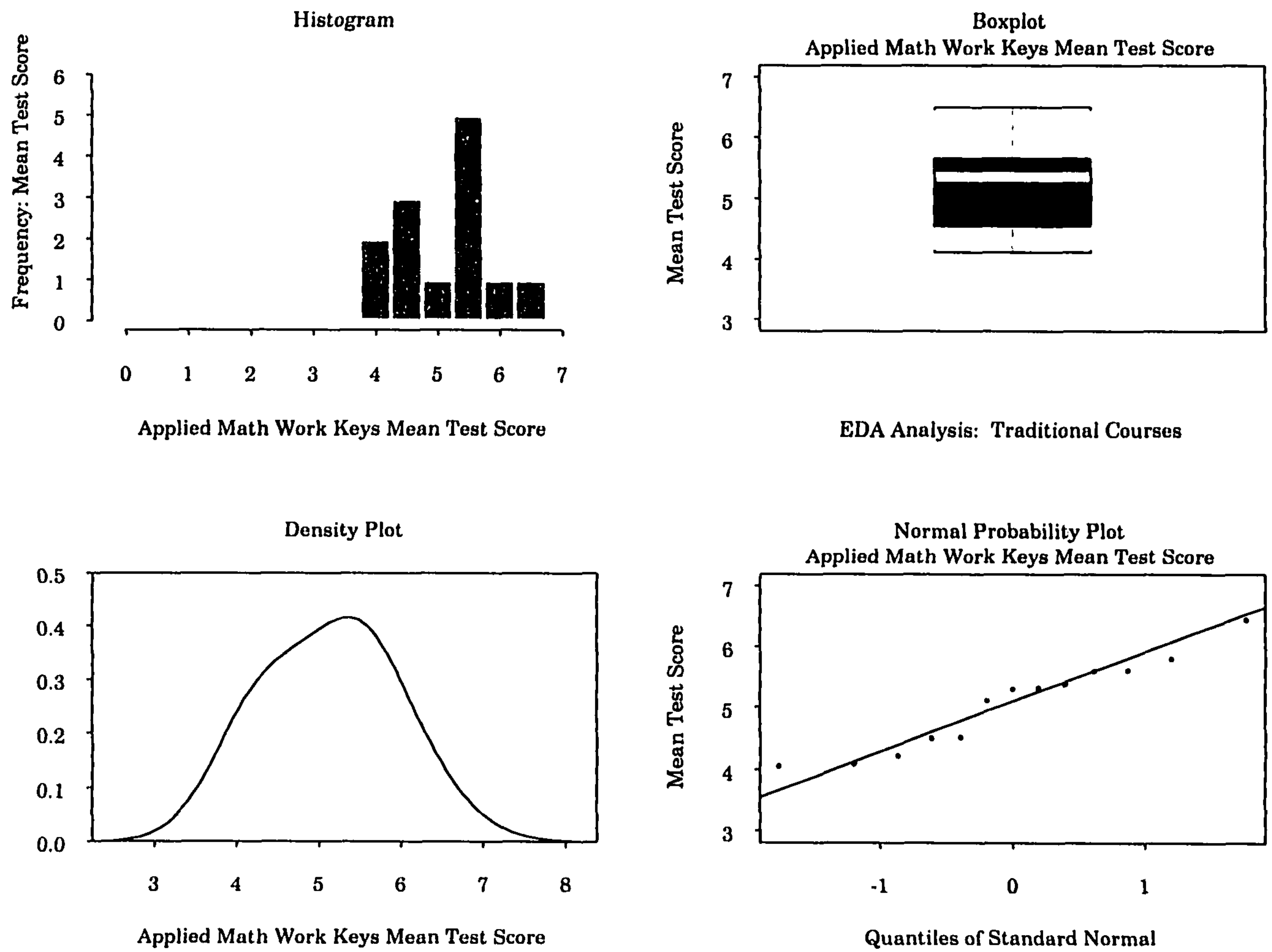


Figure 4.20. Exploratory Data Analysis plots of course x school mean Applied Math Work Keys test scores for students enrolled in traditional courses

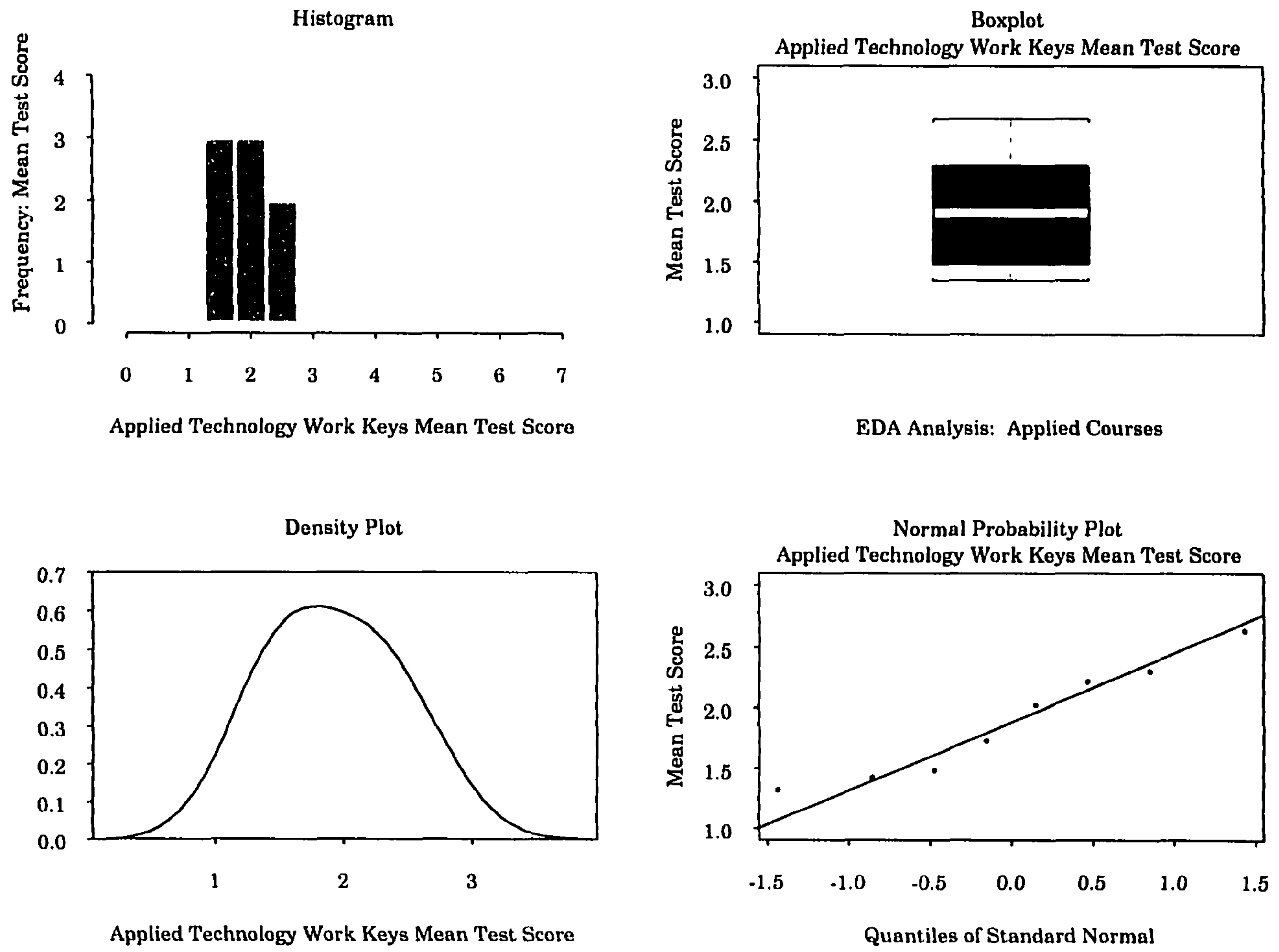


Figure 4.21. Exploratory Data Analysis plots of course x school mean Applied Technology Work Keys test scores for students enrolled in applied courses

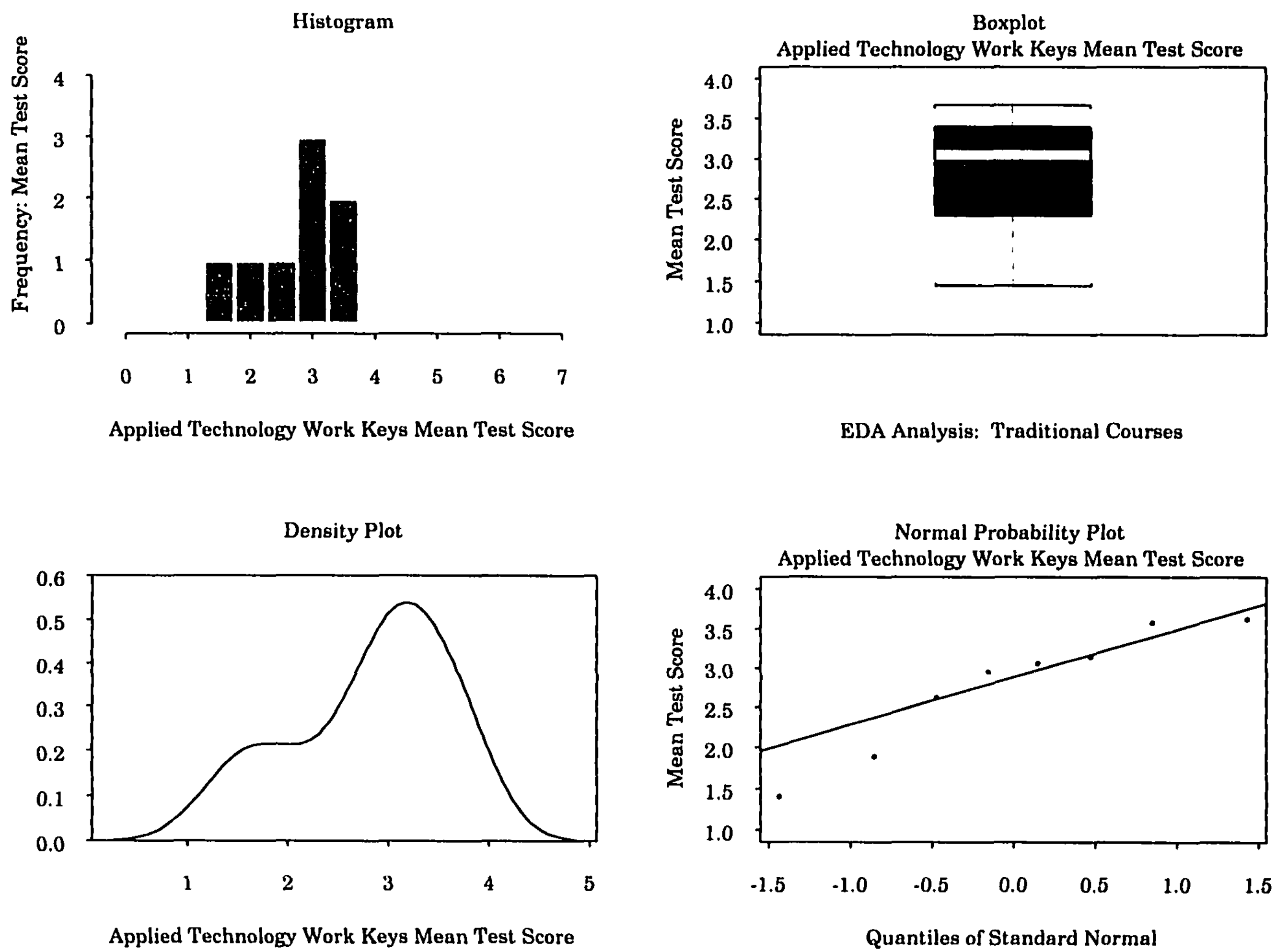


Figure 4.22. Exploratory Data Analysis plots of course x school mean Applied Technology Work Keys test scores for students enrolled in traditional courses

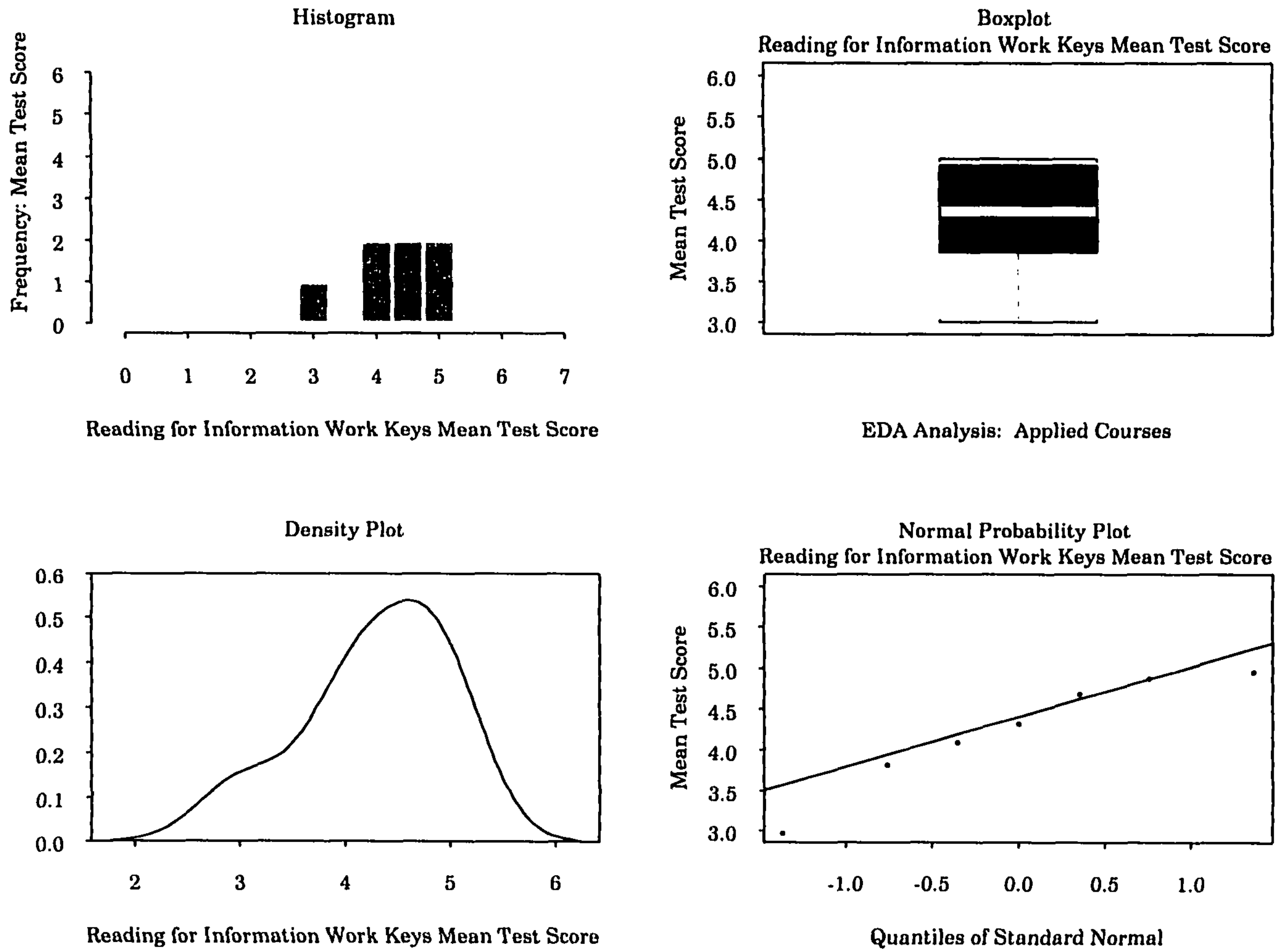


Figure 4.23. Exploratory Data Analysis plots of course x school mean Reading for Information Work Keys test scores for students enrolled in applied courses

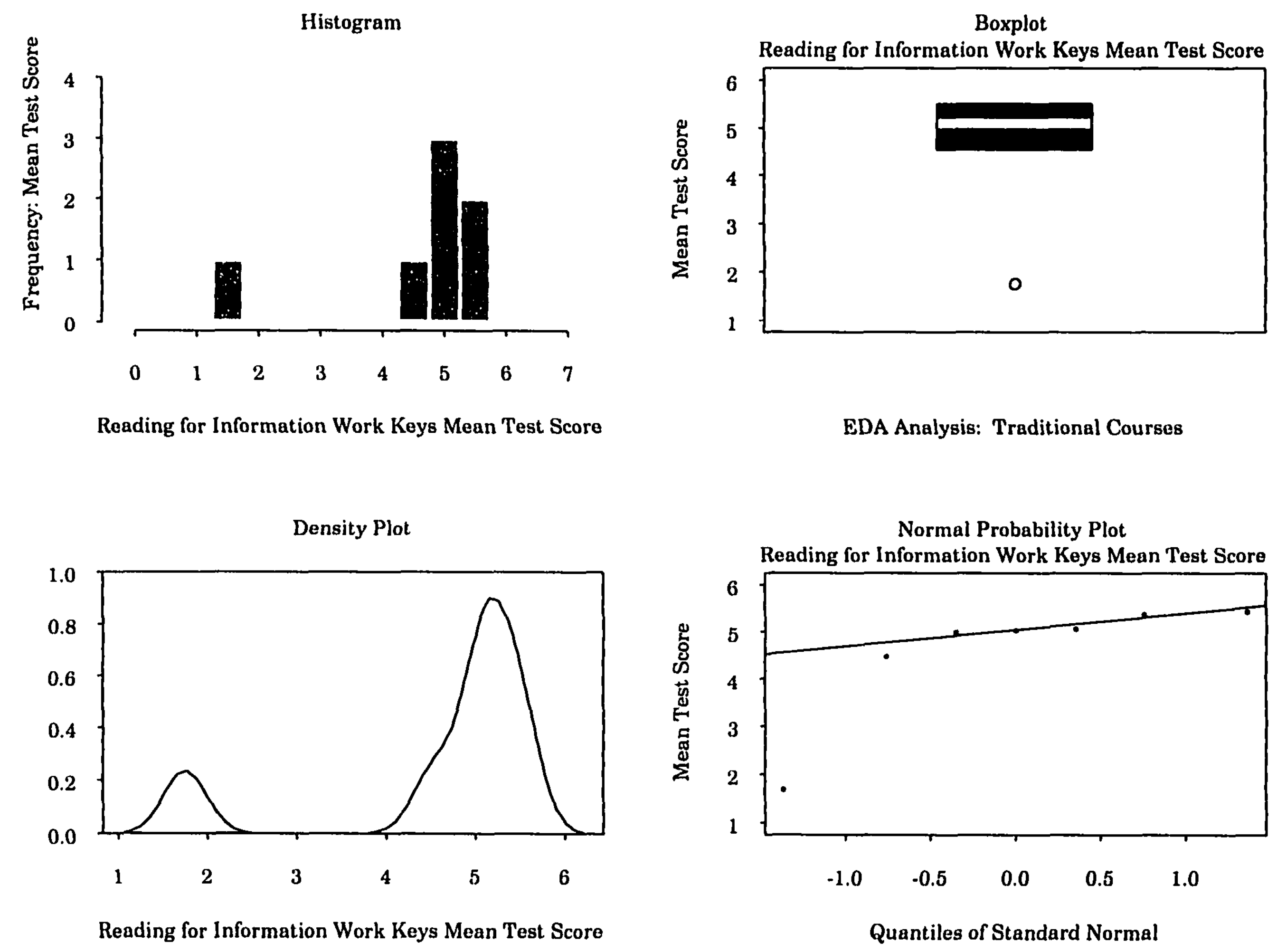
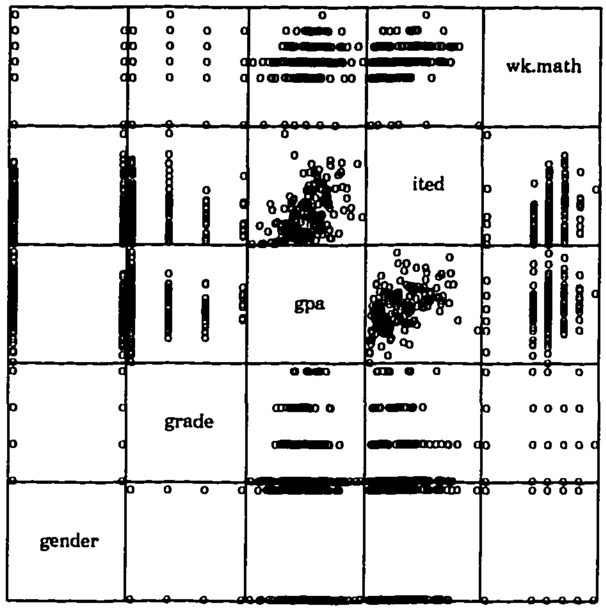


Figure 4.24. Exploratory Data Analysis plots of course x school mean Reading for Information Work Keys test scores for students enrolled in traditional courses

As a side note, the outlier shown in Figure 4.24 was checked and no evidence of erroneous data was found. The data making up the average of 1.75 were extreme--one 7 and three 0 scores--but that was not considered to be sufficient reason to discard the data; particularly since the course in which the students were enrolled was neither English nor Applied Communications.

The next series of graphs, Figures 4.25 through 4.30, are pairwise scatter plots of data used to visually evaluate the bivariate relationship of variables for subsequent HLM modeling. The data, and therefore the graphs, are divided into subgroups for presentation. Data for students taking each of the 3 Work Keys tests are placed in one of four groups: applied versus traditional courses crossed with courses that are likely to cover test concepts versus those courses that are not. For example, the top plot of Figure 4.25 contains data from students who took the Applied Math test and were enrolled in either Applied Math I or II; the bottom plot contains data from students who took the Applied Math test and were enrolled in either Algebra I or one of the Traditional Math II courses. The top plot of Figure 4.26 contains data from students who took the Applied Math test and were enrolled in Principles of Technology, Applied Communications, or Applied Biology/Chemistry courses; the bottom plot contains data from students who took the Applied Math test and were enrolled in Physics, Traditional English, or Traditional Biology/Chemistry.

Students Enrolled in Applied Math Courses



Students Enrolled in Traditional Math Courses

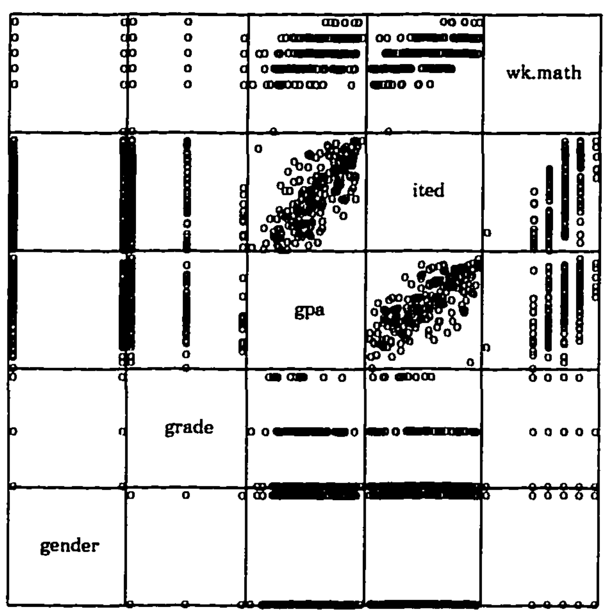
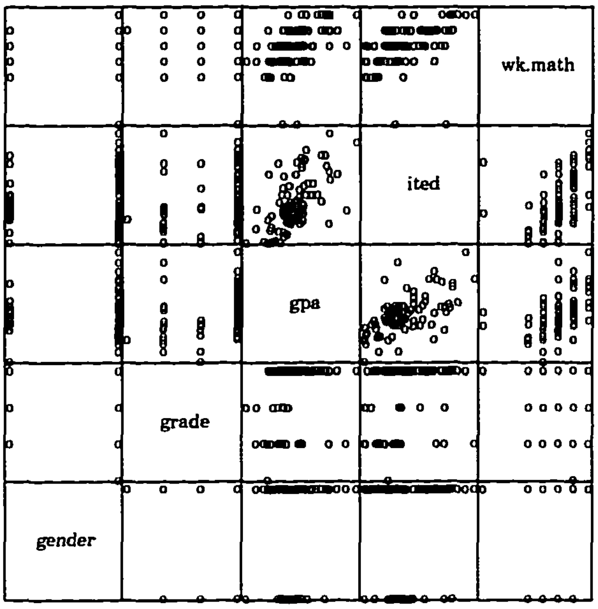


Figure 4.25. Scatterplot matrix of data collected for students enrolled in math courses and taking Applied Mathematics test

Students Enrolled in Applied Courses other than Math



Students Enrolled in Traditional Courses other than Math

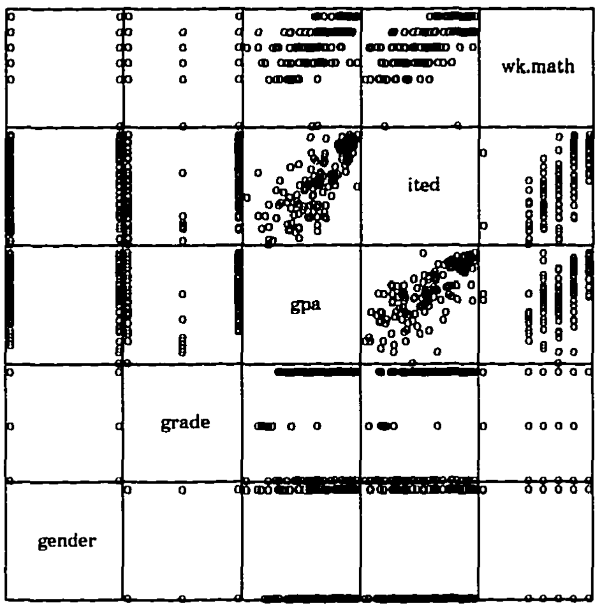
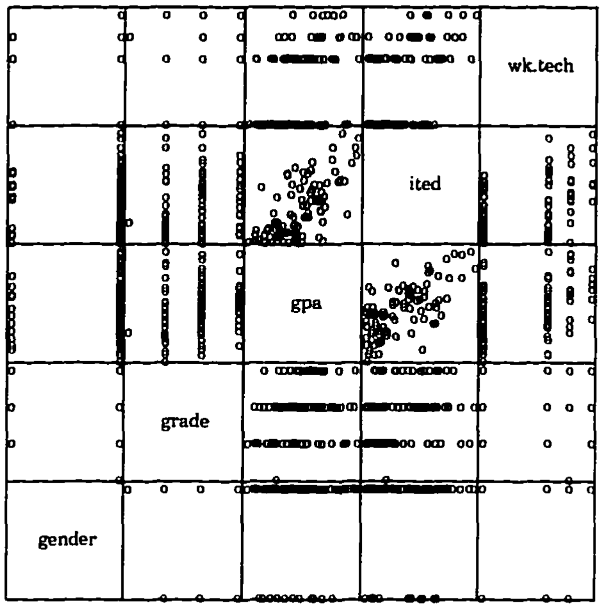


Figure 4.26. Scatterplot matrix of data collected for students enrolled in non-math courses and taking Applied Mathematics test

Students Enrolled in Principles of Technology Courses



Students Enrolled in Physics Courses

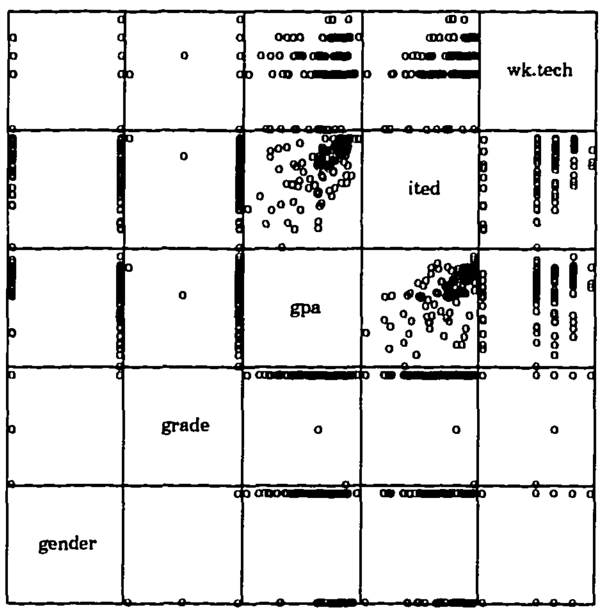
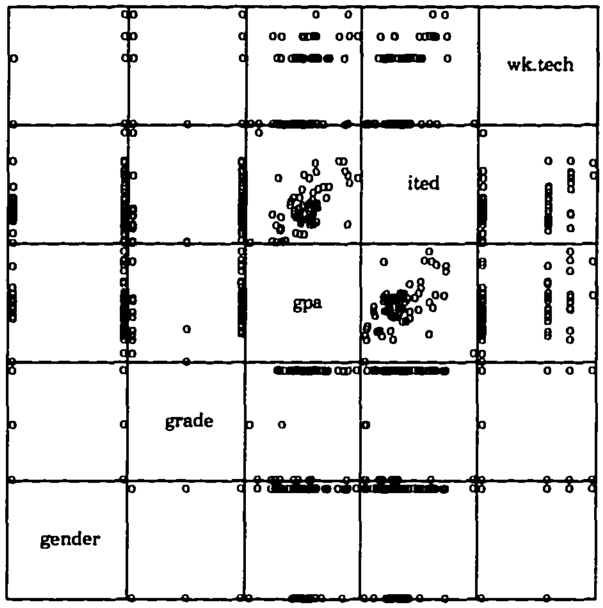


Figure 4.27. Scatterplot matrix of data collected for students enrolled in technology courses and taking Applied Technology test

Students Enrolled in Applied Courses other than Principles of Technology



Students Enrolled in Traditional Courses other than Physics

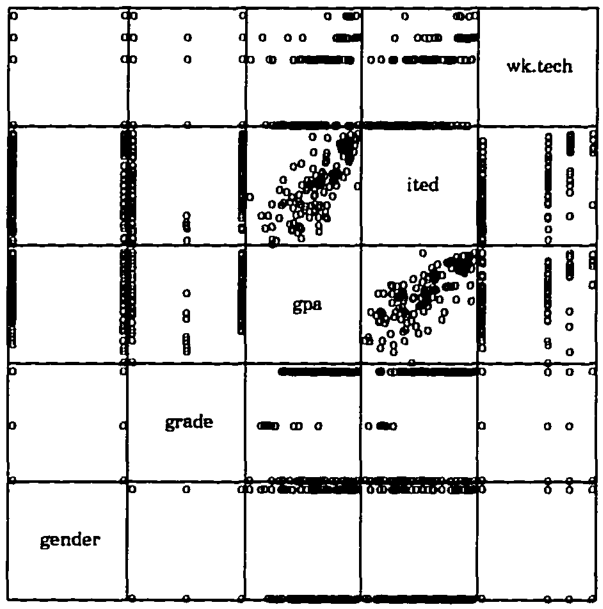
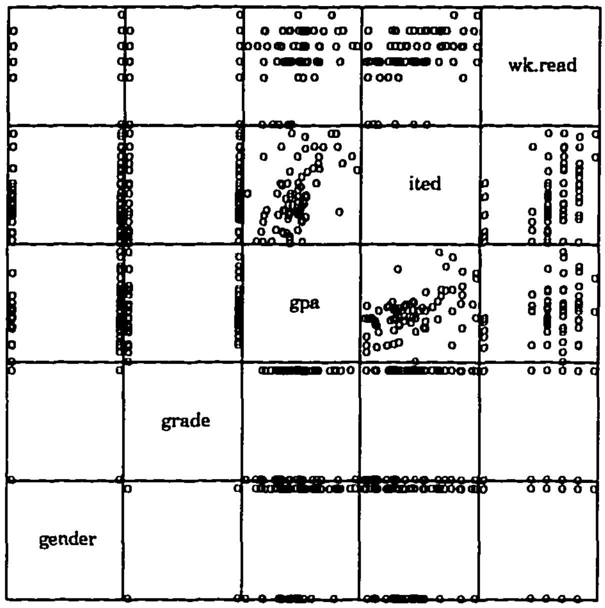


Figure 4.28. Scatterplot matrix of data collected for students enrolled in non-technology courses and taking Applied Technology test

Students Enrolled in Applied Communications Courses



Students Enrolled in Traditional English Courses

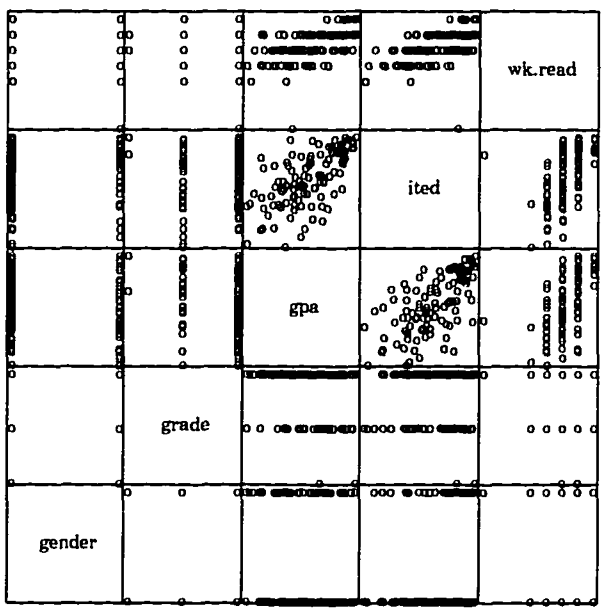
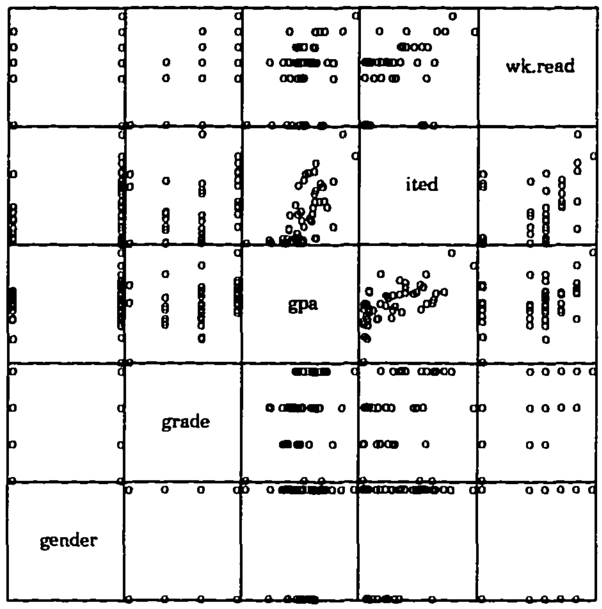


Figure 4.29. Scatterplot matrix of data collected for students enrolled in communications courses and taking Reading for Information test

Students Enrolled in Applied Courses other than Communications



Students Enrolled in Traditional Courses other than English

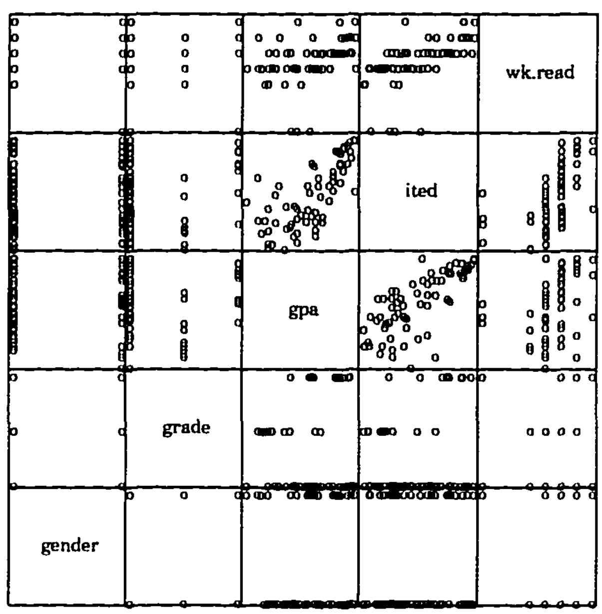


Figure 4.30. Scatterplot matrix of data collected for students enrolled in non-communications courses and taking Reading for Information test

In addition to scatter plots, a series of correlation matrices for each Work Keys test were generated. Tables 4.10 through 4.12 contain Pearson pairwise correlation coefficients for Level 1 variables data collected from students who took Work Keys tests.

Spearman correlation coefficients were also checked and compared against the Pearson coefficients, since one might question whether or not some of the data are measured on an interval or ordinal scale; however the differences were not deemed noteworthy and therefore Pearson coefficients were used.

Table 4.10. Correlation Matrix for Students taking the Applied Mathematics Work Keys test (Coefficient / (Cases) / 2-tailed Significance)

	WK.Math	ITED	GPA	Grade	Gender
WK.Math	1.000 (559) p = na	0.524 (559) p = .000	0.478 (559) p = .000	0.235 (559) p = .000	0.123 (559) p = .004
ITED	0.524 (559) p = .000	1.000 (559) p = na	0.739 (559) p = .000	0.062 (559) p = .143	-0.063 (559) p = .137
GPA	0.478 (559) p = .000	0.739 (559) p = .000	1.000 (559) p = na	0.042 (559) p = .326	-0.169 (559) p = .000
Grade	0.235 (559) p = .000	0.062 (559) p = .143	0.042 (559) p = .326	1.000 (559) p = na	0.050 (559) p = .243
Gender	0.123 (559) p = .004	-0.063 (559) p = .137	-0.169 (559) p = .000	0.050 (559) p = .243	1.000 (559) p = na

Table 4.11. Correlation Matrix for Students taking the Applied Technology Work Keys test (Coefficient / (Cases) / 2-tailed Significance)

	WK.Tech	ITED	GPA	Grade	Gender
WK.Tech	1.000	0.532	0.462	0.229	0.262
	(384)	(384)	(384)	(384)	(384)
	p = na	p = .000	p = .000	p = .000	p = .000
ITED	0.532	1.000	0.800	0.379	-0.081
	(384)	(384)	(384)	(384)	(384)
	p = .000	p = na	p = .000	p = .000	p = .114
GPA	0.462	0.800	1.000	0.336	-0.155
	(384)	(384)	(384)	(384)	(384)
	p = .000	p = .000	p = na	p = .000	p = .002
Grade	0.229	0.379	0.336	1.000	-0.041
	(384)	(384)	(384)	(384)	(384)
	p = .000	p = .000	p = .000	p = na	p = .421
Gender	0.262	-0.081	-0.155	-0.041	1.000
	(384)	(384)	(384)	(384)	(384)
	p = .000	p = .114	p = .002	p = .421	p = na

There is significant correlation between the ITED scores and GPA for all three Work Keys tests. Partial correlation coefficients for the scores of the three Work Keys tests versus GPA were calculated to see if, after controlling for ITED scores, grade, and gender, the correlation coefficient for GPA was still significant. The results were inconsistent and rather than attempting to include both GPA and ITED in the hierarchical model as separate variables, the decision was made to combine GPA and ITED scores into a new variable called "ACHIEV".

Table 4.12. Correlation Matrix for Students taking the Reading for Information Work Keys test (Coefficient / (Cases) / 2-tailed Significance)

	WK.Read	ITED	GPA	Grade	Gender
WK.Read	1.000 (290) p = na	0.541 (290) p = .000	0.479 (290) p = .000	0.225 (290) p = .000	-0.207 (290) p = .000
ITED	0.541 (290) p = .000	1.000 (290) p = na	0.759 (290) p = .000	0.151 (290) p = .010	-0.097 (290) p = .101
GPA	0.479 (290) p = .000	0.759 (290) p = .000	1.000 (290) p = na	0.076 (290) p = .196	-0.207 (290) p = .000
Grade	0.225 (290) p = .000	0.151 (290) p = .010	0.076 (290) p = .196	1.000 (290) p = na	0.046 (290) p = .433
Gender	-0.207 (290) p = .000	-0.097 (290) p = .101	-0.207 (290) p = .000	0.046 (290) p = .433	1.000 (290) p = na

The following formula is used to calculate ACHIEV:

$$\mathbf{ACHIEV} = \frac{\{\mathbf{ITED} + (\mathbf{25*GPA})\}}{\mathbf{2}}$$

The new variable is simply a linear combination of the GPA and ITED variables. GPA is multiplied by 25 to change from a 4-point scale to a 100-point scale similar to the ITED scale. By adding the two and then taking the average, the new variable is returned to a 100-point scale. This new variable has the advantage of damping the effect of a "bad day" that a student might have experienced when taking the ITED, but still allows the use of a standardized test independent of variations in grading policies. It avoids the problems inherent in using highly correlated variables together as independent variables in a regression equation, while minimizing the loss of information one would incur by eliminating one or the other. As can be seen in Table 4.13, the pairwise correlation coefficient for each of the three Work Keys tests and ACHIEV is, with one exception (0.530 versus 0.532), higher than the coefficients for any of the combinations of the Work Keys tests and either ITED or GPA. Also, when controlling for ACHIEV, Table 4.14 shows that neither GPA nor ITED have correlation coefficients significant at the 5 percent level for any of the Work Keys tests.

Table 4.13. Correlation Matrix comparing coefficients for each of the 3 Work Keys tests and the variables ACHIEV, ITED, and GPA (Coefficient / (Cases) / 2-tailed Significance)

	ACHIEV	ITED	GPA
WK.Math	0.540 (559) p = .000	0.524 (559) p = .000	0.478 (559) p = .000
WK.Tech	0.530 (384) p = .000	0.532 (384) p = .000	0.462 (384) p = .000
WK.Read	0.549 (290) p = .000	0.541 (290) p = .000	0.479 (290) p = .000

Table 4.14. Correlation coefficients for each of the 3 Work Keys tests and ITED and GPA when controlling for ACHIEV (Coefficient / (Cases) / 2-tailed Significance)

	ITED	GPA
WK.Math	0.040 (556) p = .343	-0.040 (556) p = .343
WK.Tech	0.095 (381) p = .063	-0.095 (381) p = .063
WK.Read	0.058 (287) p = .323	-0.058 (287) p = .323

Table 4.15. ACHIEV Correlation Matrix for Students taking the Applied Math Work Keys test (Coefficient / (Cases) / 2-tailed Significance)

	WK.Math	ACHIEV	Grade	Gender
WK.Math	1.000 (559) p = na	0.540 (559) p = .000	0.235 (559) p = .000	0.123 (559) p = .004
ACHIEV	0..540 (559) p = .000	1.000 (559) p = na	.057 (559) p = .178	-0.116 (559) p = .006
Grade	0.235 (559) p = .000	.057 (559) p = .178	1.000 (559) p = na	0.050 (559) p = .243
Gender	0.123 (559) p = .004	-0.116 (559) p = .006	0.050 (559) p = .243	1.000 (559) p = na

Table 4.16. ACHIEV Correlation Matrix for Students taking the Applied Technology test (Coefficient / (Cases) / 2-tailed Significance)

	WK.Tech	ACHIEV	Grade	Gender
WK.Tech	1.000 (384) p = na	0.530 (384) p = .000	0.229 (384) p = .000	0.262 (384) p = .000
ACHIEV	0.530 (384) p = .000	1.000 (384) p = na	0.380 (384) p = .000	-0.117 (384) p = .022
Grade	0.229 (384) p = .000	0.380 (384) p = .000	1.000 (384) p = na	-0.041 (384) p = .421
Gender	0.262 (384) p = .000	-0.117 (384) p = .022	-0.041 (384) p = .421	1.000 (384) p = na

Table 4.17. ACHIEV Correlation Matrix for Students taking the Reading for Information test (Coefficient / (Cases) / 2-tailed Significance)

	WK.Read	ACHIEV	Grade	Gender
WK.Read	1.000 (290) p = na	0.549 (290) p = .000	0.225 (290) p = .000	-0.207 (290) p = .000
ACHIEV	0.549 (290) p = .000	1.000 (290) p = na	0.129 (290) p = .028	-0.149 (290) p = .011
Grade	0.225 (290) p = .000	0.129 (290) p = .028	1.000 (290) p = na	0.046 (290) p = .433
Gender	-0.207 (290) p = .000	-0.149 (290) p = .011	0.046 (290) p = .433	1.000 (290) p = na

Statistical Data Analysis

As previously mentioned, this section includes two components: the first covering questions regarding academic achievement of students in applied courses versus those in traditional courses; and the second, addressing questions concerning the impact of curricula type (applied versus traditional) on employability skills. Differences in academic achievement were evaluated through the use of the Wilcoxon signed-rank test. The impact of curricula type was analyzed two ways; first using a paired sample test on *school x course* Work Keys test means, and second using Hierarchical Linear Modeling (HLM) techniques.

Academic Achievement

Table 4.18 and Table 4.19 present the paired sample data used in the GPA and ITED analyses respectively. The Wilcoxon signed-rank test was used in both cases because, as noted earlier in this chapter, the data showed evidence of a non-normal distribution. For the reader unfamiliar with the signed-rank test, an excellent overview may be found in Snedecor and Cochran (1989, pp. 140-142). In each of the tables, the data is presented first, followed by the results of the paired sample test. Sample size data is included along with the means to allow the reader a sense of the number of data points entering into the calculation of the mean.

Table 4.18. Paired GPA data analyzed by Wilcoxon signed-rank test (Sample Size refers to the number of students from which the *school x course* mean was calculated)

	Applied Students		Traditional Students	
ID Number	Sample Size	Mean GPA	Sample Size	Mean GPA
1	24	1.527	37	2.194
2	10	1.714	12	2.296
3	42	1.832	49	3.310
4	12	2.737	17	3.522
5	21	2.136	22	3.395
6	15	2.075	12	3.004
7	23	2.681	21	3.269
8	31	2.475	25	3.688
9	94	2.694	86	2.840
10	44	1.934	49	2.683
11	40	1.722	57	3.058
12	22	2.026	39	3.246
13	51	2.082	54	3.450
14	35	2.160	63	2.900
15	14	2.561	10	3.318
16	45	1.617	32	2.795
17	26	1.825	33	2.993
18	11	2.605	10	3.555
19	28	2.144	34	3.251

Wilcoxon signed-rank statistic (V) = 0

n = 19

p-value = 0

Alternative hypothesis: true mu is not equal to 0

Table 4.19. Paired ITED data analyzed by Wilcoxon signed-rank test (Sample Size refers to the number of students from which the *school x course* mean was calculated)

	Applied Students		Traditional Students	
ID Number	Sample Size	Mean ITED	Sample Size	Mean ITED
1	23	16.696	37	31.595
2	10	15.100	12	32.667
3	39	19.154	46	77.130
4	11	40.727	16	71.438
5	19	26.263	22	59.909
6	14	29.071	9	51.667
7	23	30.522	7	73.429
8	10	20.900	56	47.714
9	41	29.829	44	45.409
10	38	18.553	53	62.264
11	19	30.211	37	65.162
12	45	33.556	50	75.980
13	33	31.152	63	58.667
14	13	47.000	10	83.100
15	41	25.902	32	63.125
16	26	36.577	32	70.156
17	28	31.071	34	73.676

Wilcoxon signed-rank statistic (V) = 0

n = 17

p-value = 0

Alternative hypothesis: true mu is not equal to 0

Curricula Type

Tables 4.20 through 4.23 provide the *school x course* mean Work Keys score data and statistical test results for the covariate-free analyses.

The paired data for the Applied Technology and Reading for Information Work Keys tests were analyzed using the Wilcoxon signed-rank test due to the apparent non-normal distributions evident from the EDA plots. The points in question from the Applied Technology data--ID numbers 3 and 5 on the Traditional side, and ID number 8 on the Applied side--came from courses other than Principles of Technology or Physics. The primary point in question from the Reading for Information data--ID number 4 on the Traditional side --came from a course other than traditional English. Because of the unusual nature of this outlier and the relatively small number of data points comprising this mean, the Reading for Information data were reanalyzed with data pair ID number 4 deleted. The results of this analysis are provided in Table 4.23. The data were analyzed using both the Student's t-test and the Wilcoxon signed-rank test. This allows the reader to compare the results from the two different analytical methods on the same set of data. Note that the p-value from the Student's t-test is 0.0355, while the p-value from the Wilcoxon method is similar at 0.0312.

Table 4.20. Paired Work Keys Applied Math data analyzed by Student's t-test (Sample Size refers to the number of students from which the *school x course* mean was calculated)

	Applied Students		Traditional Students	
ID Number	Sample Size	Mean Test Score	Sample Size	Mean Test Score
1	19	4.158	35	4.114
2	9	3.667	11	4.273
3	18	3.944	18	5.167
4	22	4.409	21	4.571
5	29	4.276	24	5.667
6	93	4.376	83	4.554
7	33	3.606	41	4.146
8	13	4.077	37	5.378
9	29	4.759	44	5.432
10	12	5.083	8	6.500
11	43	3.953	31	5.355
12	26	4.769	23	5.652
13	19	5.474	28	5.857

Student's t-test statistic (t) = -5.3325

degrees of freedom = 12

p-value = 0.0002

Alternative hypothesis: true mean of differences is not equal to 0

95 percent confidence interval: (-1.0959954 to -0.4601648)

sample estimates: mean of (x - y) = -0.7780801

Table 4.21. Paired Work Keys Applied Technology data analyzed by Wilcoxon signed-rank test (Sample Size refers to the number of students from which the *school x course* mean was calculated)

	Applied Students		Traditional Students	
ID Number	Sample Size	Mean Test Score	Sample Size	Mean Test Score
1	35	1.457	48	3.104
2	12	2.250	16	3.188
3	93	1.516	84	1.452
4	35	2.057	42	3.619
5	29	1.759	45	1.933
6	12	2.667	9	3.667
7	3	2.333	8	3.000
8	20	1.350	27	2.667

Wilcoxon signed-rank statistic (V) = 1

n = 8

p-value = 0.0156

Alternative hypothesis: true mu is not equal to 0

Table 4.22. Paired Work Keys Reading for Information test data analyzed by Wilcoxon signed-rank test (Sample Size refers to the number of students from which the *school x course* mean was calculated)

	Applied Students		Traditional Students	
ID Number	Sample Size	Mean Test Score	Sample Size	Mean Test Score
1	12	3.000	10	5.100
2	92	4.130	85	4.541
3	12	5.000	27	5.444
4	7	3.857	4	1.750
5	28	4.357	46	5.130
6	12	4.917	8	5.500
7	18	4.722	28	5.071

Wilcoxon signed-rank statistic (V) = 7

n = 7

p-value = 0.2969

Alternative hypothesis: true mu is not equal to 0

Table 4.23. Paired Work Keys Reading for Information test data (minus sample 4) analyzed by both Wilcoxon signed-rank test and Student's t-test (Sample Size refers to the number of students from which the *school x course* mean was calculated)

	Applied Students		Traditional Students	
ID Number	Sample Size	Mean Test Score	Sample Size	Mean Test Score
1	12	3.000	10	5.100
2	92	4.130	85	4.541
3	12	5.000	27	5.444
5	28	4.357	46	5.130
6	12	4.917	8	5.500
7	18	4.722	28	5.071

Wilcoxon signed-rank statistic (V) = 0

n = 6

p-value = 0.0312

Alternative hypothesis: true mu is not equal to 0

Student's t-test statistic (t) = -2.8585

degrees of freedom = 5

p-value = 0.0355

Alternative hypothesis: true mean of differences is not equal to 0

95 percent confidence interval: (-1.475 to -0.078)

sample estimates: mean of (x - y) = -0.777

The rest of this section is devoted to the analysis using a Hierarchical Linear Model (HLM). The data analysis using the HLM for Windows program involved three stages:

1. Building a sufficient statistics matrix (SSM) file from the raw data
2. Fitting models using information in the SSM file
3. Evaluating the fitted models through residuals analyses

The numbers included in the final set of data used to develop the SSM file for each of the three Work Keys tests are shown in Table 4.24. The slight difference in student and class totals between Table 4.24 and Tables 4.3 and 4.4 are due to the removal of data series that did not match up across levels.

The analyses started with a fully unconditional model; that is, outcomes at each level were modeled as a mean plus a random error, with no predictor variables included. This models allowed a determination as to how variations in the outcomes were allocated across the three levels. The fully unconditional model used for all three Work Keys data sets follows Table 4.24.

Table 4.24. Number of Level 1, Level 2, and Level 3 units in the HLM analyses of the three Work Keys tests

		Applied Mathematics	Applied Technology	Reading for Information
Level 1	Students	590	381	290
Level 2	Classes	71	43	43
Level 3	Schools	8	6	5

Level-1 Unconditional Model. Within each classroom, we model student employability skills (that is, Work Keys assessment test scores) as a function of a classroom mean plus a random student-level error:

$$Y_{(ijk)} = \pi_{0(jk)} + e_{(ijk)}, \text{ where}$$

$Y_{(ijk)}$ is the Work Keys score of student *i* in classroom *j* and school *k*.

$\pi_{0(jk)}$ is the mean Work Keys test score in classroom *j* and school *k*.

$e_{(ijk)}$ is a Level-1 random effect that represents the deviation of student *ijk*'s score from the classroom mean score. These residual effects are assumed normally distributed with a mean of 0 and a variance of σ^2.

Level-2 Unconditional Model. Each classroom mean Work Keys test score, $\pi_{0(j)}$, in the above Level-1 model can be viewed as an outcome varying randomly around some school mean Work Keys test score:

$$\pi_{0(jk)} = \beta_{00(k)} + r_{0(jk)}, \text{ where}$$

$\beta_{00(k)}$ is the mean Work Keys test score in school *k*.

$r_{0(jk)}$ is a Level-2 random effect that represents the deviation of classroom *jk*'s score from the school mean score. These residual effects are assumed normally distributed with a mean of 0 and a variance of τ_π. Also, intra-classroom variability for each of the k schools is assumed the same.

Level-3 Unconditional Model. Each school mean Work Keys test score, $\beta_{00(k)}$, in the above Level-2 model can be viewed as an outcome varying randomly around some grand mean Work Keys test score:

$$\beta_{00(k)} = \gamma_{000} + u_{00(k)}, \text{ where}$$

γ_{000} is the grand mean Work Keys test score

$u_{00(k)}$ is a Level-3 random effect that represents the deviation of school *k*'s score from the grand mean score. These residual effects are assumed normally distributed with mean of 0 and variance of τ_β.

Bryk and Raudenbush (1992) provide formulas for use in partitioning variance and calculating reliabilities for this three-level model:

> This simple three-level model partitions the total variability in the outcome $Y_{(ijk)}$ into its three components: (Level 1) among [students] within classrooms, σ^2; (Level 2) among classrooms within schools, τ_π; and (Level 3) among schools, τ_β. It also allows us to estimate the proportion of variation that is within classrooms, among classrooms within schools, and among schools. That is,
> $\sigma^2/(\sigma^2 + \tau_\pi + \tau_\beta)$ is the proportion of variance within classrooms; [8.4]
> $\tau_\pi/(\sigma^2 + \tau_\pi + \tau_\beta)$ is the proportion of variance among classrooms within schools; and [8.5]
> $\tau_\beta/(\sigma^2 + \tau_\pi + \tau_\beta)$ is the proportion of variance among schools [8.6].
> (p. 177)

Tables 4.25 through 4.27 provide results of the fully unconditional three-level analysis of the Applied Mathematics, Applied Technology, and Reading for Information Work Keys tests.

The maximum likelihood point estimate for the grand mean Applied Math Work Keys score, as seen in Table 4.25, is 4.79 with a standard error of 0.17, yielding the following 95% confidence interval:

$$4.79 \pm 1.96\ (0.17) = (4.46, 5.12).$$

The variance components data are interpreted as follows: At the school level, 0.164 is the estimated variability of the true school means around the grand mean. The *df*, χ^2, and *p value* columns provide information regarding a test of whether or not this estimated value is significantly greater than zero. If it is not, then one may reasonably assume that all schools have the same mean.

The school variance component for the Applied Technology test data yields a *p value* of 0.115. Based on this result, a 2-Level model was used for this data analysis, rather than the 3-Level model used for the analyses of the Applied Mathematics and Reading for Information data.

At the class level the variance is the estimated variability of the true class means around the school mean. The variance calculated from the deviation of each student's Work Keys score from the classroom mean is given as the Level 1 variance component. In all three Work Keys tests, most of the variation in the outcome is at the student level. The variance decomposition is remarkably similar between the Applied Mathematics and Reading for Information tests, with 18% to 19% at the class level and a little over 9% at the school level. This is somewhat less than the typical between-school variance suggested by Bryk and

Raudenbush (1992). They state that, "... about 14% of the variance in initial status was between schools, which is consistent with results typically encountered in cross-sectional studies of school effects" (p. 7).

The maximum likelihood point estimate for the grand mean Applied Technology Work Keys score is 2.31 with a standard error of 0.17, yielding a 95% confidence interval below the minimum competency level determined by ACT:

$$2.31 \pm 1.96\ (0.17) = (1.98,\ 2.64).$$

The 95% confidence interval for Reading for Information is:

$$4.28 \pm 1.96\ (0.28) = (3.73,\ 4.83).$$

Table 4.25. Three-level, fully unconditional model: Applied Mathematics

Fixed Effect	*Coefficient*	*se*	*t* ratio
School (i.e., grand) mean, γ_{000}	4.791	0.172	27.902

Random Effect	*Variance Component*	*df*	χ^2	*p value*
Level 1 (Students), $e_{(ijk)}$	1.249			
Level 2 (Classes), $r_{0(jk)}$	0.311	63	130.811	0.000
Level 3 (Schools), $u_{00(k)}$	0.164	7	29.059	0.000

Variance Decomposition (by level)		
Level 1	Students	72.4%
Level 2	Classes	18.1%
Level 3	Schools	9.5%

Table 4.26. Three-level, fully unconditional model: Applied Technology

Fixed Effect	*Coefficient*	*se*	*t* ratio
School (i.e., grand) mean, γ_{000}	2.309	0.172	13.460

Random Effect	*Variance Component*	*df*	χ^2	*p value*
Level 1 (Students), $e_{(ijk)}$	2.955			
Level 2 (Classes), $r_{0(jk)}$	0.482	37	96.403	0.000
Level 3 (Schools), $u_{00(k)}$	0.048	5	8.843	0.115

Variance Decomposition (by level)		
Level 1	Students	84.8%
Level 2	Classes	13.8%
Level 3	Schools	1.4%

Table 4.27. Two-level, fully unconditional model: Applied Technology (5 classes and 10 students deleted from data based on residual analysis--see discussion of Applied Technology residual analysis later in chapter)

Fixed Effect	*Coefficient*	*se*	*t* ratio
School (i.e., grand) mean, γ_{000}	2.336	0.152	15.347

Random Effect	*Variance Component*	*df*	χ^2	*p value*
Level 1 (Students), $e_{(ijk)}$	2.977			
Level 2 (Classes), $r_{0(jk)}$	0.535	37	104.848	0.000

Variance Decomposition (by level)		
Level 1	Students	84.8%
Level 2	Classes	15.2%

Table 4.28. Three-level, fully unconditional model: Reading for Information

Fixed Effect	*Coefficient*	*se*	*t* ratio
School (i.e., grand) mean, γ_{000}	4.280	0.277	15.435

Random Effect	*Variance Component*	*df*	χ^2	*p value*
Level 1 (Students), $e_{(ijk)}$	1.996			
Level 2 (Classes), $r_{0(jk)}$	0.540	38	85.046	0.000
Level 3 (Schools), $u_{00(k)}$	0.253	4	16.493	0.003

Variance Decomposition (by level)		
Level 1	Students	71.6%
Level 2	Classes	19.3%
Level 3	Schools	9.1%

Bryk and Raudenbush (1992) also provide formulas to calculate the reliability of the least squares estimated coefficients for classrooms (Level 2) and schools (Level 3) of the three-level model.

> For each classroom jk at Level 2,
>
> $$\text{reliability}\left(\hat{\pi}_{0(jk)}\right) = \tau_\pi / \left[\tau_\pi + \sigma^2 / n_{jk}\right] \qquad [8.7]$$
>
> is the reliability of a classroom sample mean for use in discrimination among classrooms within the same school. For any school k at Level 3,
>
> $$\text{reliability}\left(\hat{\beta}_{00k}\right) = \frac{\tau_\beta}{\tau_\beta + \left\{\sum\left[\tau_\pi + \sigma^2/n_{jk}\right]^{-1}\right\}^{-1}} \qquad [8.8]$$
>
> is the reliability of the school's sample mean as an estimate of its true mean.
>
> The averages of these reliabilities across classrooms (Equation 8.7) and schools (Equation 8.8) may be viewed as summary measures of the reliability of the class and school means, respectively. (pp. 177-178)

Table 4.29 provides the results of the Level 2 and Level 3 reliability calculations for each of the three Work Keys tests.

One should interpret the reliability as "the ratio of the true score or parameter variance, relative to the observed score or total variance of the sample mean" (Bryk & Raudenbush, 1992, p. 40). As one can gather from the formulas, the reliability will be close to 1 when the parameter variance component is much larger than the error variance component. This will occur when one of two conditions hold: (a) the group means vary substantially across the units of the level in question with constant sample size per group; or (b) the sample sizes are large (Bryk & Raudenbush, 1992, p. 40). A reliability close to 0.5 indicates that the parameter variance component and the error variance component are essentially equal. The Level 2 reliabilities shown in Table 4.29 indicate that it would be somewhat difficult to discriminate among classrooms within the same school simply by looking at the classroom sample mean; the estimates of the within- and between-class variances are approximately the same size.

Table 4.29. Level 2 and Level 3 reliability calculations for each of the three Work Keys tests

		Applied Mathematics	Applied Technology	Reading for Information
Level 2	Classes	0.589	0.545	0.533
Level 3	Schools	0.739	0.518	0.788

The next step in the HLM process was to begin introducing predictor variables for consideration in a model. Bryk and Raudenbush (1992) state:

> The substantive theory under study should suggest a relatively small number of predictors for possible consideration in the Level-1 model. There are two questions here: (a) Should a candidate [a_p] be included in the model? If yes, (b) how should its coefficient be specified: random, fixed, or nonrandomly varying?
>
> Initially, the Level-2 predictors are held aside and the analysis focuses on comparing some alternative hierarchical models, each of which is unconditional at Level 2. (p. 201)

Bryk and Raudenbush (1992) provide some guidelines on page 202 as to how one can address these two questions. They suggest that one look to see if the fixed effect of a_p is significant as a method of answering part (a), and to investigate whether or not there is evidence of slope heterogeneity, that is, whether or not Var(π_p) > 0, to provide some indication of whether the predictor should be specified as random, fixed, nonrandomly varying. Bryk and Raudenbush (1992) go on to say that:

> Statistical evidence of slope heterogeneity includes the point estimates [of the variance of π_p] and the corresponding homogeneity test statistics (χ^2 and likelihood-ratio tests ...). Also useful in this regard are the estimated reliabilities for the OLS intercepts and slopes.
>
> When the reliabilities become small (e.g., < 0.05), the variances we wish to estimate are likely to be close to zero .... Inspection of the reliabilities may suggest that a random Level-1 coefficient be respecified as either fixed or nonrandomly varying. (p. 202)

There were three predictor variables evaluated at Level 1 of the model. They were ACHIEV (a composite of ITED and GPA), GENDER, and GRADE (the student's year in school). There were three predictor variables evaluated at Level

2 of the model. They were TYPE (applied versus traditional course), RELVNT (a dummy variable used to indicate whether or not the course material was relevant to the material covered in the Work Keys test--for example, math courses were relevant to the Applied Math test, while communications courses were not), and CMACHIEV (the class mean value of ACHIEV). No school level predictor variables were used; however, school-level random components were included in all versions of the Applied Math and Reading for Information models.

Two tables are presented for each of the Work Keys tests; the first is the model including the three predictor variables at Levels 1 and 2. This allows the reader to see the fixed effect coefficients and *p values* for all variables under consideration. The second model includes only those predictor variables yielding significant coefficients. The base model used to evaluate the Work Keys test data sets, including descriptions of all potential variables and coefficients, precedes the tables. In the case of the Applied Technology analysis, Level-3 of the model was eliminated, but was otherwise identical to the model for the other two Work Keys analyses. A considerable number of alternate models were evaluated but not included here. For the interested reader, Appendix D contains several examples of preliminary models and the reports generated by the HLM software during the Applied Math model-building process. Several residuals diagnostic plots were included after the tables for each of the Work Keys tests. These plots were used to check assumptions of normality and homogeneous variances.

Level-1 Model:

$$Y_{(ijk)} = \pi_{0(jk)} + \pi_{1(jk)}a_{1(ijk)} + \pi_{2(jk)}a_{2(ijk)} + \pi_{3(jk)}a_{3(ijk)} + e_{(ijk)}$$

Where

$Y_{(ijk)}$ is the Work Keys test score of student *i* in class *j* and school *k*.

$\pi_{0(jk)}$ is the mean Work Keys score of 9th grade females with a class average ACHIEV score in class *j* and school *k*.

$\pi_{1(jk)}$ is the predicted change to mean Work Keys score in class *j* and school *k* when the student is a male. This is a "gender-gap" coefficient.

$a_{1(ijk)}$ is a dummy variable associated with student gender. The coding is 0 for a female student and 1 for a male student.

$\pi_{2(jk)}$ is the predicted change to mean Work Keys score in class *j* and school *k* as a result of the student's grade level (9th, 10th, 11th, or 12th)

$a_{2(ijk)}$ is a dummy variable associated with student grade level. The coding is 0 for a student in 9th grade, 1 for a student in 10th grade, 2 for a student in 11th grade, and 3 for a student in 12th grade.

$\pi_{3(jk)}$ is the predicted change to mean Work Keys score in classroom *j* and school *k* per unit change in the student's class-centered ACHIEV score.

$a_{3(ijk)}$ is the class-centered ACHIEV score of student *i* in class *j* and school *k*.

$e_{(ijk)}$ is a Level-1 random effect that represents the deviation of student *ijk*'s score from the predicted score. These residual effects are assumed normally distributed with a mean of 0 and a variance of σ^2.

Level-2 Model:

$$\pi_{0(jk)} = \beta_{00(k)} + \beta_{01(k)}X_{1(jk)} + \beta_{02(k)}X_{2(jk)} + \beta_{03(k)}X_{3(jk)} + r_{0(jk)}$$

$$\pi_{1(jk)} = \beta_{10(k)}$$

$$\pi_{2(jk)} = \beta_{20(k)}$$

$$\pi_{3(jk)} = \beta_{30(k)}$$

Where

$\beta_{00(k)}$ is the mean Work Keys test score of 9th grade females in applied non-math courses with a school mean ACHIEV score in school *k*.

$\beta_{01(k)}$ is the predicted change to overall class mean Work Keys test score of 9th grade females in non-math courses with a school mean ACHIEV score in school *k* when traditional curricula are used rather than applied curricula. This is a "curricula-gap" coefficient.

$X_{1(jk)}$ is a variable associated with curriculum type used in classroom *j* in school *k*. The coding is 0 for an applied and 1 for a traditional course.

$\beta_{02(k)}$ is the predicted change to overall class mean Work Keys test score of 9th grade females in applied courses with a school mean ACHIEV score in school *k* when the applied course is a math course rather than a non-math course. This is a "relevant course" coefficient.

$X_{2(jk)}$ is a dummy variable used to identify whether or not a course is "relevant" to the Work Keys test taken in school *k*. The coding is 0 for a non-relevant course and 1 for a relevant course.

$\beta_{03(k)}$ is the predicted change to overall class mean Work Keys test score of 9th grade females in non-math applied courses in school *k* per unit change in the student's school mean-centered ACHIEV score.

$X_{3(jk)}$ is the student's school mean-centered ACHIEV score.

$r_{0(jk)}$ is a Level-2 random effect that represents the deviation of class *j*'s Level-1 intercept coefficient from its predicted value based on the Level-2 model. The random effects in Level 2 equations are assumed to be correlated. They are also assumed to be multivariate normal with a mean of 0. The variance of this effect is designated as τ_{π}.

$\beta_{10(k)}$ is the mean slope, averaged across classes within school *k*, relating student gender to Work Keys score. When the coefficient is considered a fixed effect, as it is here with $\pi_{1(jk)}$ assumed equal to $\beta_{10(k)}$, it implies that there are not statistically significant differences in the relationship between a student's gender and the Work Keys test score from class to class within a school.

$\beta_{20(k)}$ is the mean slope, averaged across classes within school *k*, relating student grade to Work Keys score.

$\beta_{30(k)}$ is the mean slope, averaged across classes within school *k*, relating student class-centered ACHIEV score to Work Keys score for school *k*.

Level-3 Model:

$$\beta_{00(k)} = \gamma_{000} + u_{00(k)}$$

$$\beta_{01(k)} = \gamma_{010}$$

$$\beta_{02(k)} = \gamma_{020}$$

$$\beta_{03(k)} = \gamma_{030}$$

$$\beta_{10(k)} = \gamma_{100}$$

$$\beta_{20(k)} = \gamma_{200}$$

$$\beta_{30(k)} = \gamma_{300}$$

Where

γ_{000} is the grand mean Work Keys test score of 9th grade females, with class-centered student ACHIEV scores equal to 0, in applied non-math classes where the school-centered class mean ACHIEV score is also equal to 0.

$u_{00(k)}$ is a Level-3 random effect that represents the deviation of school *k*'s mean Work Keys score from the grand mean value based on the Level-3 model. The random effects in Level 3 equations are assumed to be correlated. They are also assumed to be multivariate normal with a mean of 0. The variance of this effect is designated as τ_β.

γ_{010} is the curricula gap coefficient averaged over schools. When the coefficient is considered a fixed effect, as it is here with $\beta_{10(k)}$ assumed equal to γ_{010}, it implies that there are not statistically significant differences in the relationship between the curricula type and the Work Keys test score from school to school.

γ_{020} is the mean slope, averaged over schools, relating the impact of "relevant" courses on the Work Keys test score.

γ_{030} is the mean slope, averaged over schools, relating mean class school-centered ACHIEV score to (Work Keys) test score.

γ_{100} is the mean slope, averaged over schools, relating gender to test score.

γ_{200} is the mean slope, averaged over schools, relating grade to test score.

γ_{300} is the mean slope, averaged over schools, relating student class-centered ACHIEV score to test score.

Table 4.30 provides information on the impact of incorporating all six predictor variables in the model; three at Level-1 and three at Level-2. Class mean ACHIEV and grade level were not significant at the 5% level in this model and both were eliminated from the subsequent model. The significance level of each of the two predictor variables was checked in the absence of the other variable to guard against mistakenly deleting a significant variable.

Approximately 64% of the class-level variation is explained by the model used to generate Table 4.30. This is worth noting so that the class-level variation may be compared with that in Table 4.31 based on a model with fewer terms.

Table 4.30. HLM estimates for Applied Mathematics data--all variables

Fixed Effect	*Coefficient*	*se*	*t* ratio	*p value*
Grand mean, γ_{000}	4.305	0.277	15.555	0.000
Curricula gap, γ_{010}	0.796	0.133	5.971	0.000
Relevant course, γ_{020}	-0.553	0.234	-2.366	0.018
Class mean ACHIEV, γ_{030}	-0.0002	0.002	-0.103	0.919
Gender gap, γ_{100}	0.422	0.088	4.782	0.000
Grade level, γ_{200}	0.112	0.080	1.411	0.158
Student ACHIEV, γ_{300}	0.031	0.003	10.839	0.000

Random Effect		*Variance Component*	*df*	χ^2	*p value*
Level 1 (Students),	$e_{(ijk)}$	1.014			
Level 2 (Classes),	$r_{0(jk)}$	0.111	60	98.147	0.002
Level 3 (Schools),	$u_{00(k)}$	0.134	7	40.267	0.000

Variance Reduction (by level) from Unconditional Model

Level 1	Students	18.8%
Level 2	Classes	64.3%

Table 4.31. HLM estimates for Applied Mathematics data--significant variables

Fixed Effect	*Coefficient*	*se*	*t* ratio	*p value*
Grand mean, γ_{000}	4.563	0.226	20.145	0.000
Curricula gap, γ_{010}	0.776	0.130	5.960	0.000
Relevant course, γ_{020}	-0.695	0.226	-3.069	0.003
Gender gap, γ_{100}	0.417	0.088	4.734	0.000
Student ACHIEV, γ_{200}	0.030	0.003	10.833	0.000

Random Effect		*Variance Component*	*df*	χ^2	*p value*
Level 1 (Students),	$e_{(ijk)}$	1.009			
Level 2 (Classes),	$r_{0(jk)}$	0.125	61	95.217	0.004
Level 3 (Schools),	$u_{00(k)}$	0.173	7	46.184	0.000

Variance Reduction (by level) from Unconditional Model

Level 1	Students	19.2%
Level 2	Classes	59.8%

Approximately 60% of the class-level variation is explained by the above model. Noting the *p value* of 0.004 for the random effect at Level 2, however, one may conclude that a significant amount of unexplained variation remains at this level. The "Curricula gap" coefficient is a positive 0.776 indicating that students enrolled in traditional courses score a little more than 3/4 of a point higher than students enrolled in applied courses. Male students scored on average 4/10 of a point higher than female students. The ACHIEV coefficient is positive, which simply means that students with higher combined GPA and ITED scores, also scored higher on the Work Keys test than those with lower combined GPA and ITED scores. The one result that was somewhat disconcerting was the negative "relevant course" coefficient. One might realistically expect students enrolled in math courses to do better on a math test than students enrolled in non-math courses. A negative coefficient for this variable indicates the opposite is true (for this sample of students). As a check on this result, histograms of the Applied Math Work Keys test scores for students enrolled in math versus non-math courses were generated. These histograms are shown in Figure 4.31. After a rough graphical comparison of the two distributions, one could conclude that a negative coefficient is a plausible result. The mean score of the students enrolled in the math courses does appear to be just slightly above 4, while the other distribution appears to be centered at 5. Without engaging in speculation as to what caused the negative coefficient, it should be noted that students participating in this study as a result of their enrollment in English or physics, for example, could also have been concurrently enrolled in math courses.

There were no concerns resulting from an examination of the Level 2 residuals plots shown in Figure 4.32. The residuals appear to be normally distributed with mean 0 and no outliers. The Level 3 residuals plot, Figure 4.33, shows one borderline point on the normal probability plot, but with the limited number of data points (schools) in the sample and the fact that it does not show up as an outlier on the boxplot, it is not considered to be a problem.

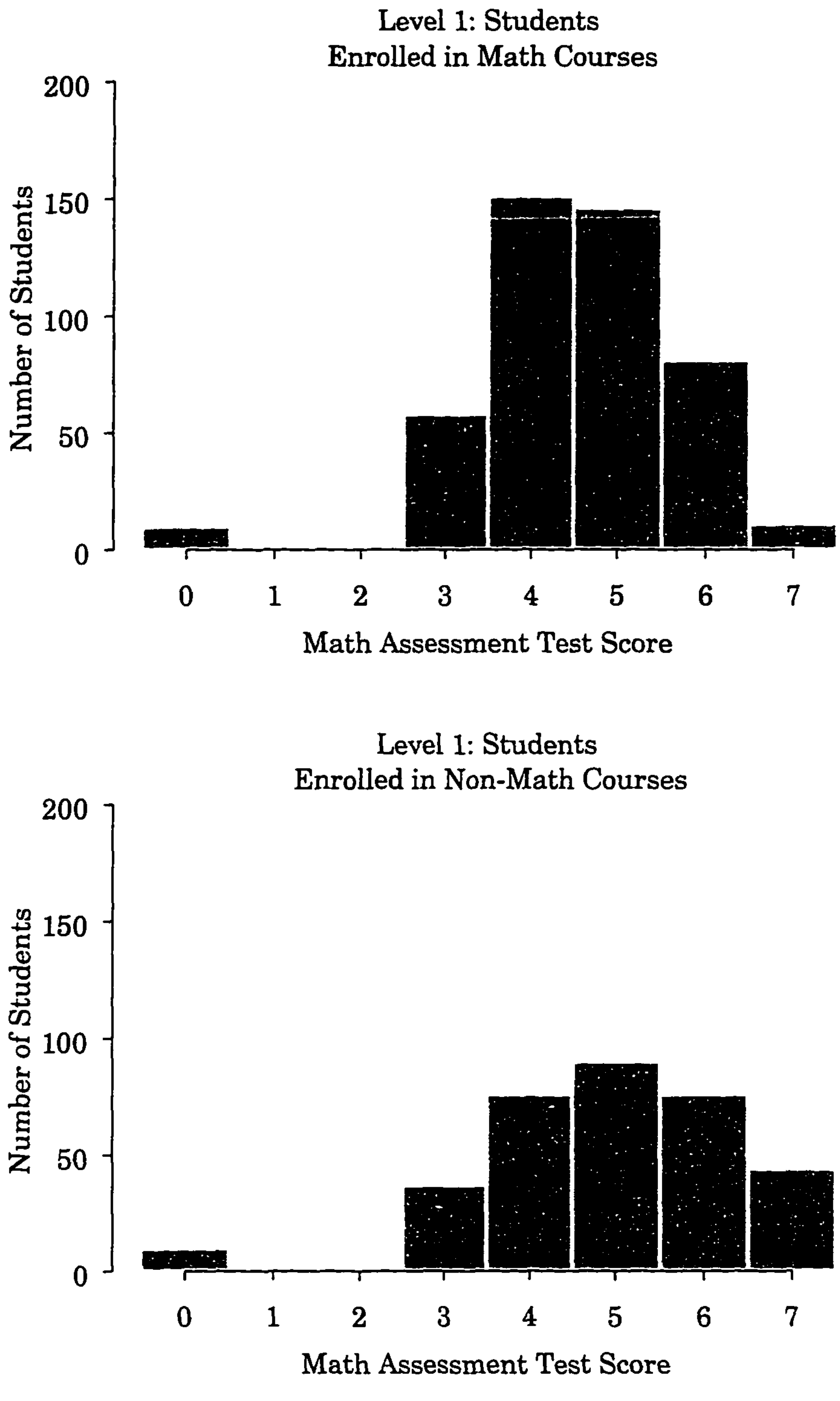


Figure 4.31. Histogram to investigate the impact of "relevant" versus "non-relevant" courses on Work Keys Applied Math test scores

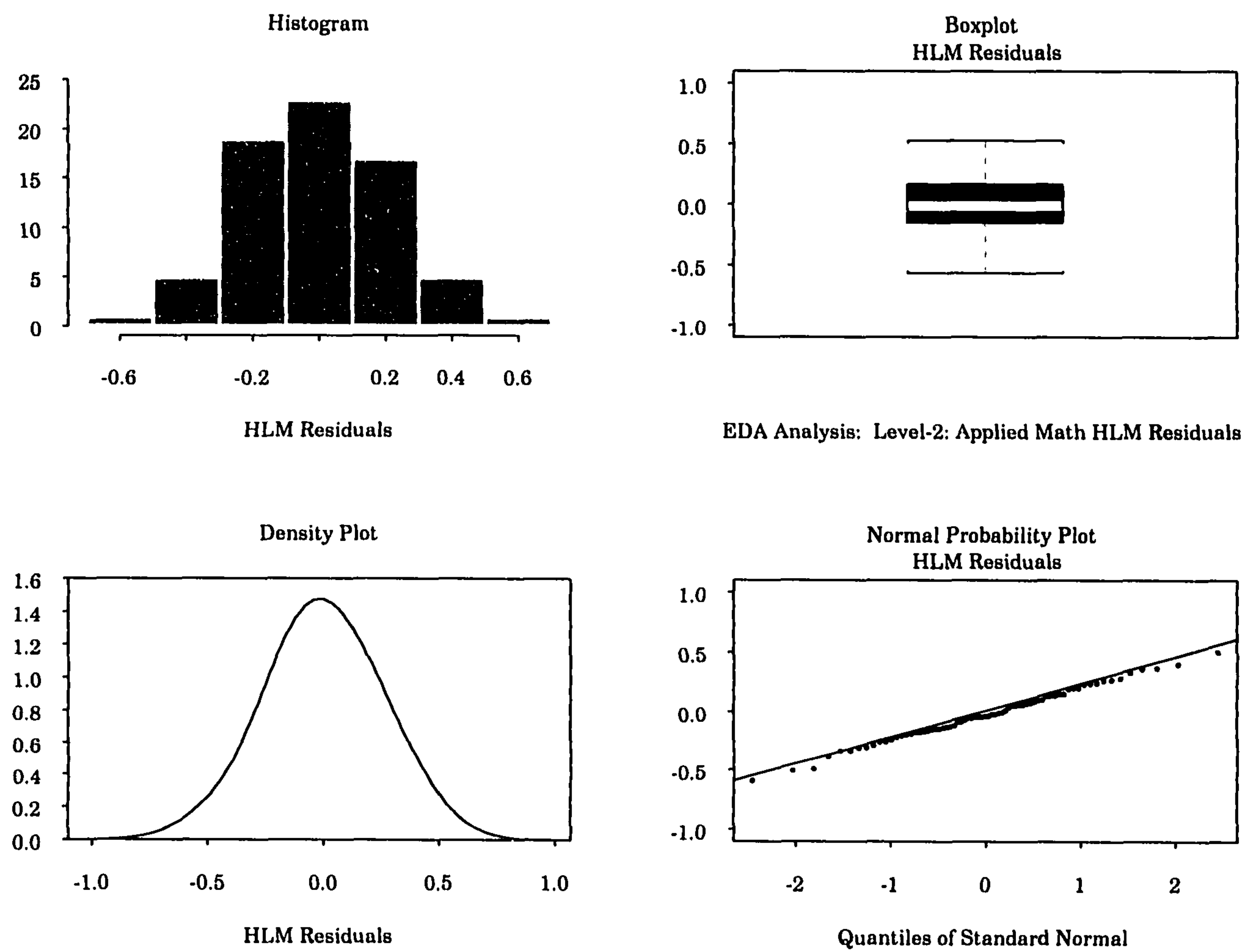


Figure 4.32. Exploratory Data Analysis plots of Applied Mathematics HLM Level 2 Residuals

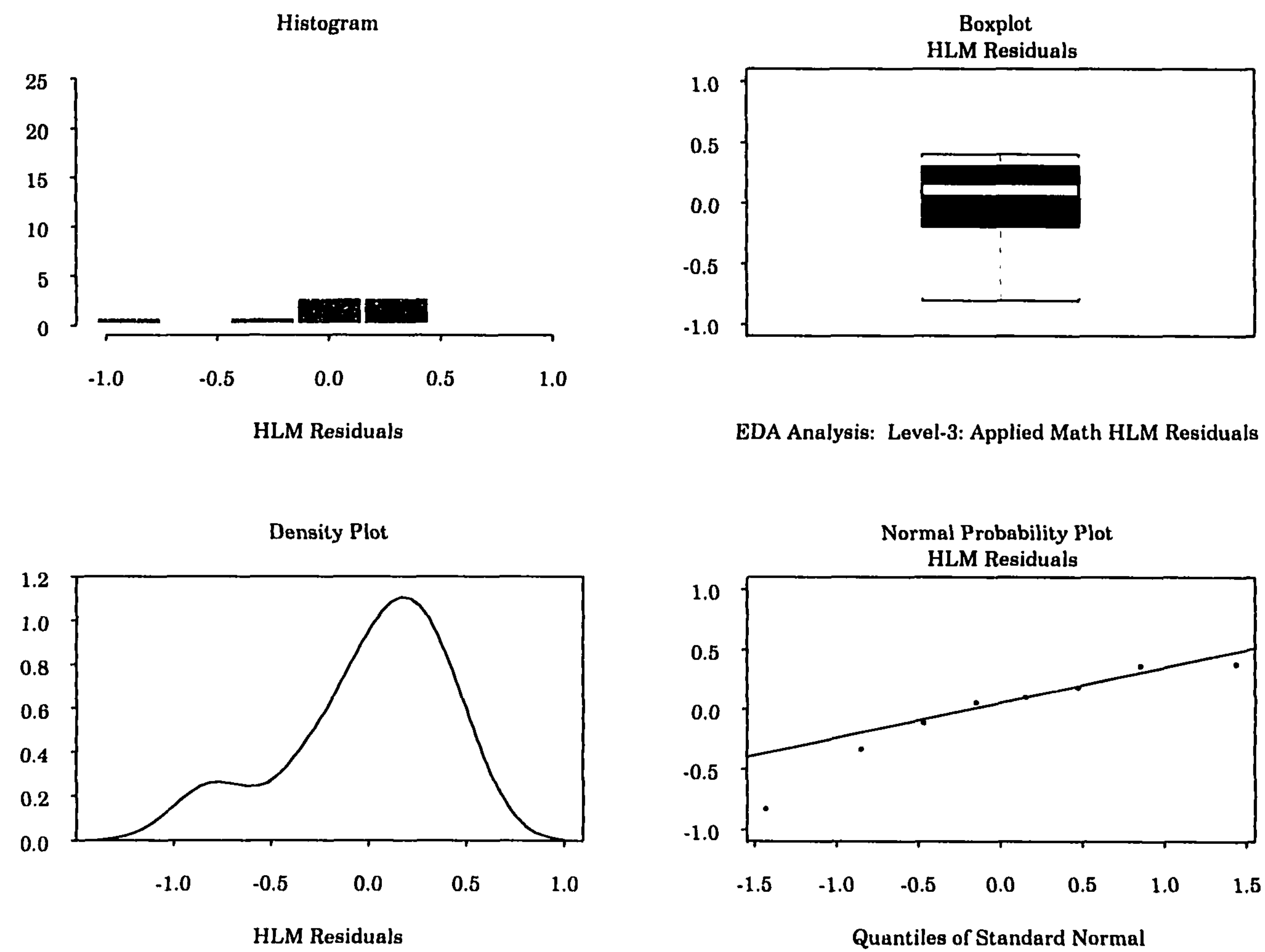


Figure 4.33. Exploratory Data Analysis plots of Applied Mathematics HLM Level 3 Residuals

Table 4.32. HLM estimates for Applied Technology data--all variables

Fixed Effect	*Coefficient*	*se*	*t* ratio	*p value*
Grand mean, γ_{00}	1.796	0.493	3.645	0.001
Curricula gap, γ_{01}	-0.739	0.384	-1.925	0.061
Relevant course, γ_{02}	0.195	0.190	1.031	0.309
Class mean ACHIEV, γ_{03}	0.060	0.012	4.951	0.000
Gender gap, γ_{10}	1.099	0.171	6.417	0.000
Grade level, γ_{20}	-0.019	0.126	-0.148	0.884
Student ACHIEV, γ_{30}	0.045	0.005	9.239	0.000

Random Effect	*Variance Component*	*df*	χ^2	*p value*
Level 1 (Students), $r_{(ij)}$	2.185			
Level 2 (Classes), $u_{0(j)}$	0.0007	39	32.927	>0.500

Variance Reduction (by level) from Unconditional Model

Level 1	Students	26.1%
Level 2	Classes	99.9%

A first pass review of the significance levels seems to indicate that the variables associated with curricula type, course relevance, and student grade level could be deleted from the model. However, once the variables associated with course relevance and student grade level are removed, as shown in Table 4.33, the type-of-curricula variable (applied versus traditional) shows up as significant at the 5% level. One can also see that there is a gender gap associated with the technology test; males on average score over one point higher than females. Also, students high on the ACHIEV scale score better on average on the Work Keys test than do those students low on the ACHIEV scale. The class mean ACHIEV level seems to have an effect on individual scores; in other words, there seems to be a tangible benefit with respect to the test score by being in a class with high mean academic achievement. The curricula gap seems to be reversed from that in the Applied Math analysis; students enrolled in applied courses did better on average than did traditional course students. One should not draw conclusions from this model however; it is presented as an example.

Table 4.33. HLM estimates for Applied Technology data--significant variables with "outlier" classes included

Fixed Effect	*Coefficient*	*se*	*t* ratio	*p value*
Grand mean, γ_{00}	1.876	0.221	8.500	0.000
Curricula gap, γ_{01}	-0.857	0.298	-2.873	0.007
Class mean ACHIEV, γ_{02}	0.063	0.008	7.983	0.000
Gender gap, γ_{10}	1.151	0.165	6.984	0.000
Student ACHIEV, γ_{20}	0.045	0.005	9.269	0.000

Random Effect	*Variance Component*	*df*	χ^2	*p value*
Level 1 (Students), $r_{(ij)}$	2.181			
Level 2 (Classes), $u_{0(j)}$	0.0007	40	34.329	>0.500

Variance Reduction (by level) from Unconditional Model		
Level 1	Students	26.2%
Level 2	Classes	99.9%

At this point the residual plots of the Level-2 data were examined (see Figure 4.34). At least two outliers showed up on the boxplot. Questionable points were also evident on the lower tail of the normal probability plot. Data for the five most extreme classes were examined. In four of these classes, only one or two full data series were available for analysis. The remaining class had four full data series available; but the test scores were extreme--one 3 and three 0s. All five of these classes were deleted from the data set and the model coefficients recalculated to gauge the impact of the classes on the coefficients.

Table 4.34. HLM estimates for Applied Technology data--significant variables with "outlier" classes deleted

Fixed Effect	*Coefficient*	*se*	*t* ratio	*p value*
Grand mean, γ_{00}	2.005	0.234	8.545	0.000
Curricula gap, γ_{01}	-0.791	0.301	-2.628	0.013
Class mean ACHIEV, γ_{02}	0.063	0.008	7.826	0.000
Gender gap, γ_{10}	1.154	0.166	6.974	0.000
Student ACHIEV, γ_{20}	0.045	0.005	9.312	0.000

Random Effect	*Variance Component*	*df*	χ^2	*p value*
Level 1 (Students), $r_{(ij)}$	2.151			
Level 2 (Classes), $u_{0(j)}$	0.0005	35	21.928	>0.500

Variance Reduction (by level) from Unconditional Model		
Level 1	Students	27.7%
Level 2	Classes	99.9%

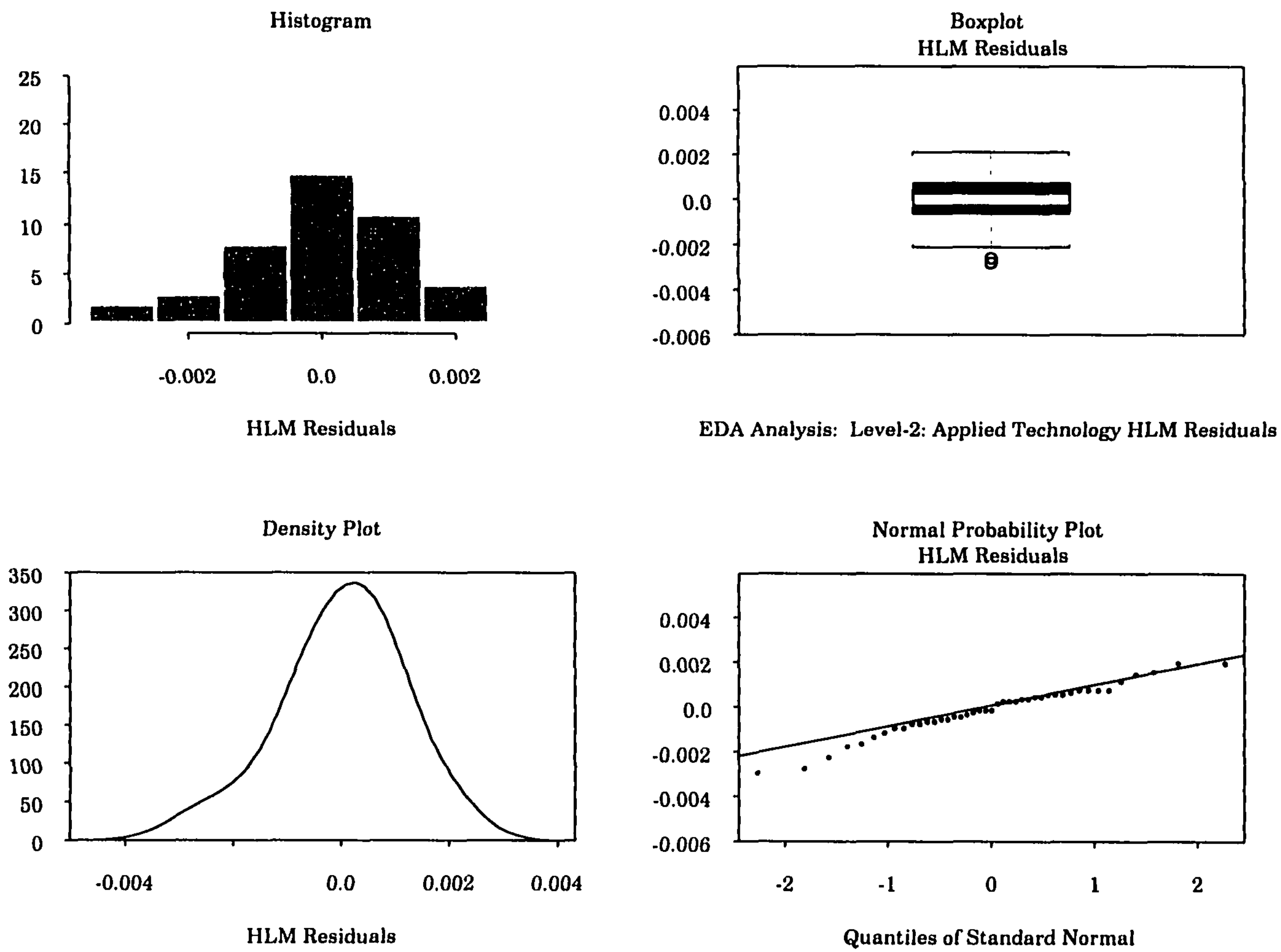


Figure 4.34. Exploratory Data Analysis plots of Applied Technology HLM Level 2 Residuals (CMACHIEV included in model)

During the evaluation of the latest set of coefficients it became obvious that the Class mean ACHIEV variable was correlated to other variables. Table 4.35 contains correlation coefficients for the three Level-2 variables. Note that the Class mean ACHIEV variable is abbreviated as CMACHIEV, while Relevant course and Curricula gap coefficients are abbreviated as RELVNT and TYPE.

Table 4.35. CMACHIEV Correlation Matrix for Students taking the Applied Technology test (Coefficient / (Cases) / 2-tailed Significance)

	CMACHIEV	TYPE	RELVNT	*RELVNT
CMACHIEV	1.000 (38) p = na	0.838 (38) p = .000	0.043 (38) p = .800	0.329 (35) p = .047
TYPE	0.838 (38) p = .000	1.000 (38) p = na	-0.161 (38) p = .334	*partial when controlling for TYPE
RELVNT	0.043 (38) p = .800	-0.161 (38) p = .334	1.000 (38) p = na	

Table 4.36 completes the coefficient analysis of the Applied Technology data. This set of coefficient tables was included to allow the reader to follow the process of analysis and to see the effect of various changes that were made to the data set and to the HLM model itself. As one can see in Table 4.36, eliminating the use of CMACHIEV results in significant positive curricula gap, relevant course, and grade level coefficients. This is in sharp contrast to the values of the

Table 4.36. HLM estimates for Applied Technology data--significant variables with "outlier" classes and CMACHIEV deleted

Fixed Effect	*Coefficient*	*se*	*t* ratio	*p value*
Grand mean, γ_{00}	-0.345	0.305	-1.133	0.266
Curricula gap, γ_{01}	1.035	0.190	5.446	0.000
Relevant course, γ_{02}	0.665	0.185	3.600	0.001
Gender gap, γ_{10}	1.119	0.178	6.304	0.000
Grade level, γ_{20}	0.430	0.106	4.071	0.000
Student ACHIEV, γ_{30}	0.045	0.005	9.306	0.000

Random Effect	*Variance Component*	*df*	χ^2	*p value*
Level 1 (Students), $r_{(ij)}$	2.223			
Level 2 (Classes), $u_{0(j)}$	0.053	35	39.530	0.274

Variance Reduction (by level) from Unconditional Model

Level 1	Students	25.3%
Level 2	Classes	90.1%

coefficients presented in Table 4.35. The curricula gap coefficient has changed signs and the grade and relevant course variables, not present in the previous model, are now significant. The intercept in the model used to generate Table 4.36 is no longer significantly different from 0 at the 5% level. The changes in the results of the analysis are classic symptoms that independent variables being considered for the model are highly correlated among themselves. As Neter et al. (1990) note:

Indications of the presence of serious multicollinearity are given by the following informal diagnostics:

1. Large changes in the estimated regression coefficients when a variable is added or deleted, or when an observation is altered or deleted.
2. Nonsignificant results in individual tests on the regression coefficients for important independent variables.
3. Estimated regression coefficients with an algebraic sign that is the opposite of that expected from theoretical considerations or prior experience.
4. Large coefficients of simple correlation between pairs of independent variables in the correlation matrix $\mathbf{r}_{xx}$.
5. Wide confidence intervals for the regression coefficients representing important independent variables. (pp. 407-408)

Neter et al. (1990) sum up the issue with the following statement:

The important conclusion we must draw is: When independent variables are correlated, the regression coefficient of any independent variable depends on which other independent variables are included in the model and which ones are left out. Thus, a regression coefficient does not reflect any inherent effect of the particular independent variable on the dependent variable but only a marginal or partial effect, given whatever other correlated independent variables are included in the model. (p. 301)

Figure 4.35 contains residuals analysis plots for the final Applied Technology model with outlier data from the original model eliminated. The residual plots show outliers based on the latest model as well. These outliers were examined and nothing unusual, other than the extremes of the test scores were noted. The outlier class on the high side had a student sample size of 11, only 2 of which scored 0s on the test. The three classes on the low side had a combined 34 zero scores out of a possible 42. Bryk and Raudenbush (1992) provide some guidance in dealing with concerns regarding normality and outliers in Chapter 9 titled “Assessing Hierarchical Models”. They state that:

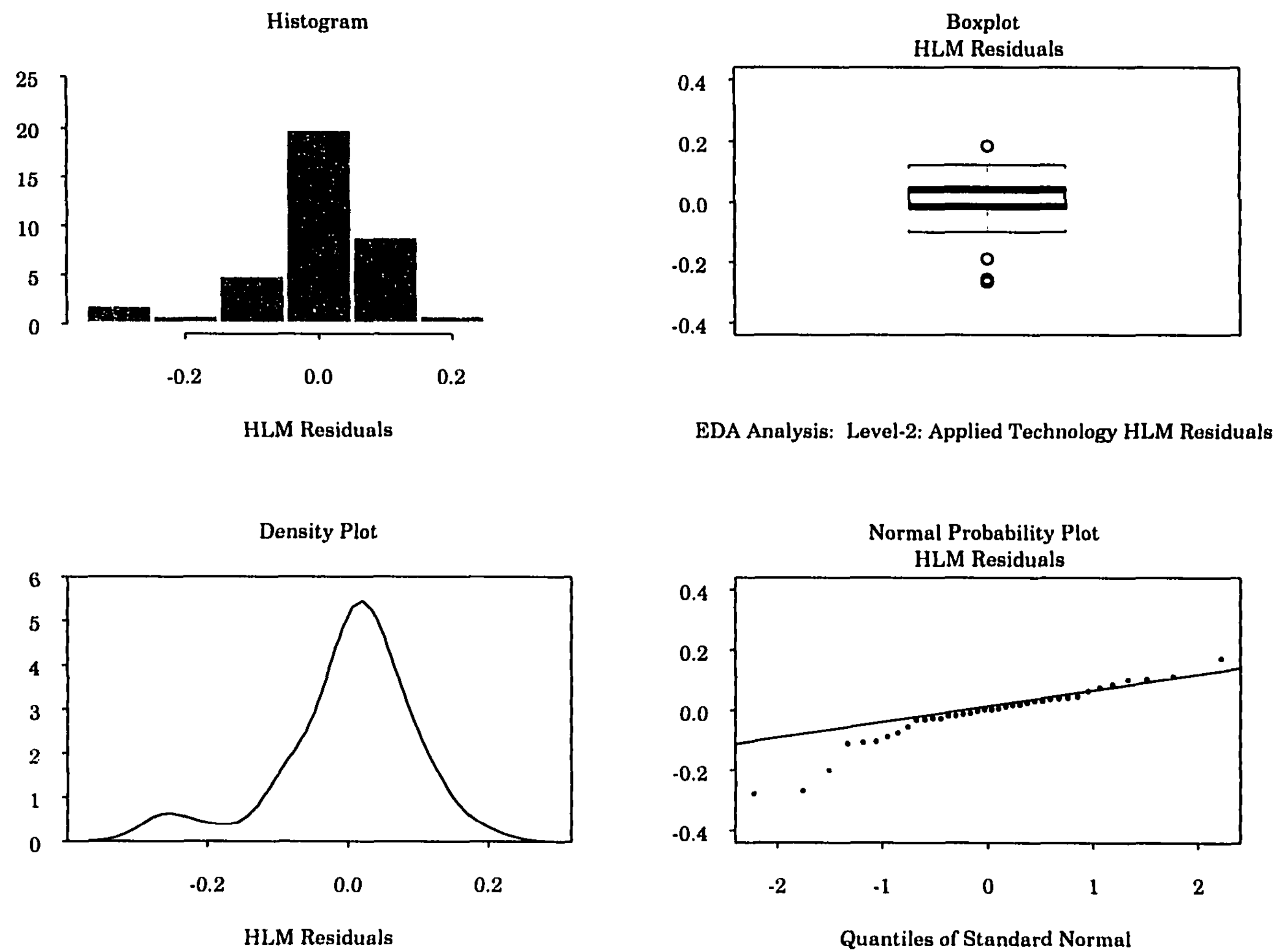


Figure 4.35. Exploratory Data Analysis plots of Applied Technology HLM Level 2 Residuals (CMACHIEV deleted)

> Estimation of the fixed effects will not be biased by a failure of the normality assumption at Level 2. However, if the Level-2 random effects have heavy tails, inferences based on normality may be sensitive to outliers. A failure of the normality assumption will affect the validity of the confidence intervals and hypothesis tests for the fixed effects. The nature of these effects depends upon the true shape of the distribution of the random effects. (p. 218)

Bryk and Raudenbush (1992) also reference a doctoral dissertation by M. Seltzer (1990) out of the University of Chicago for procedures that can be helpful if serious non-normality is encountered at Level 2. The work is titled: "The use of data augmentation in fitting hierarchical models to educational data". Such procedures were beyond the scope of this research and no attempt was made to employ them on the Applied Technology data.

The final Work Keys data analyzed by the use of HLM were the Reading for Information data. The results of the first pass analysis (all variables), are shown in Table 4.37; while the results from the analysis including only those variables significant at the 5% level are provided in Table 4.38. The curricula gap is consistent with the results from the other two Work Keys tests--students enrolled in traditional courses scored higher on average than those enrolled in applied courses. The gap between traditional and applied students was over 4/5 of a point. The gender gap was present here as well; however in this case the sign of the coefficient was reversed from that seen when the Applied Math and Technology data were analyzed. Females scored on average 1/2 of a point higher than males. Grade and the ACHIEV variable were both positive indicating that

Table 4.37. HLM estimates for Reading for Information data--all variables

Fixed Effect	*Coefficient*	*se*	*t* ratio	*p value*
Grand mean, γ_{000}	2.635	0.368	7.156	0.000
Curricula gap, γ_{010}	0.760	0.248	3.058	0.005
Relevant course, γ_{020}	0.365	0.290	1.257	0.217
Class mean ACHIEV, γ_{030}	0.001	0.001	0.628	0.534
Gender gap, γ_{100}	-0.491	0.159	-3.085	0.002
Grade level, γ_{200}	0.632	0.165	3.827	0.000
Student ACHIEV, γ_{300}	0.045	0.005	9.481	0.000

Random Effect	*Variance Component*	*df*	χ^2	*p value*
Level 1 (Students), $e_{(ijk)}$	1.439			
Level 2 (Classes), $r_{0(jk)}$	0.295	35	101.871	0.000
Level 3 (Schools), $u_{00(k)}$	0.001	4	6.497	0.164

Variance Reduction (by level) from Unconditional Model

Level 1	Students	27.9%
Level 2	Classes	45.4%

an academically gifted 12th grader did better on average than a less gifted 9th grader. These variables accounted for over 42% of the variance at the class level. A significant portion of the Level-2 variance still remains to be explained however; the χ^2 test yields a highly significant result with a *p value* of 0.000 for the Level-2 variance component. The "relevant course" variable was not a factor, either positive or negative, in this data set.

Table 4.38. HLM estimates for Reading for Information data--significant variables

Fixed Effect	*Coefficient*	*se*	*t* ratio	*p value*
Grand mean, γ_{000}	2.606	0.366	7.129	0.000
Curricula gap, γ_{010}	0.826	0.246	3.354	0.002
Gender gap, γ_{100}	-0.504	0.159	-3.175	0.002
Grade, γ_{200}	0.706	0.139	5.085	0.000
Student ACHIEV, γ_{300}	0.045	0.005	9.561	0.000

Random Effect		*Variance Component*	*df*	χ^2	*p value*
Level 1 (Students),	$e_{(ijk)}$	1.442			
Level 2 (Classes),	$r_{0(jk)}$	0.311	37	104.057	0.000
Level 3 (Schools),	$u_{00(k)}$	0.0002	4	4.251	0.373

Variance Reduction (by level) from Unconditional Model

Level 1	Students	27.8%
Level 2	Classes	42.4%

The plots for the Level 2 and Level 3 Reading for Information residuals are shown in Figures 4.36 and 4.37, respectively. These plots were reassuringly unremarkable--no outliers and no evidence of serious non-normality. The one minor concern here was the paucity of data points at Level 3. The lack of data was certainly evident in the histogram, making it of limited value as a diagnostic tool for the residuals. The limited number of data points was not a problem here since Level 3 was not a focus of this investigation.

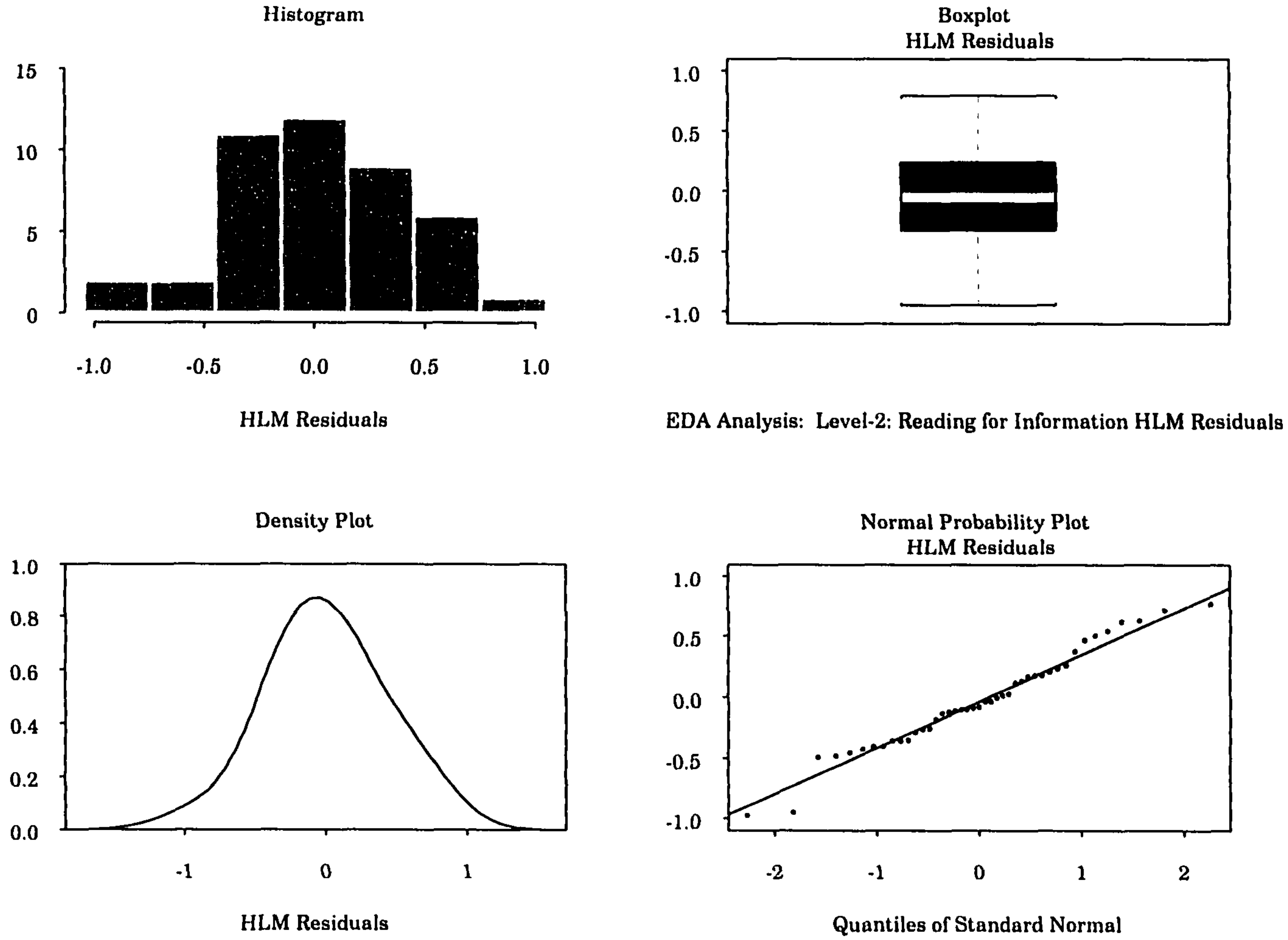


Figure 4.36. Exploratory Data Analysis plots of Reading for Information HLM Level 2 Residuals

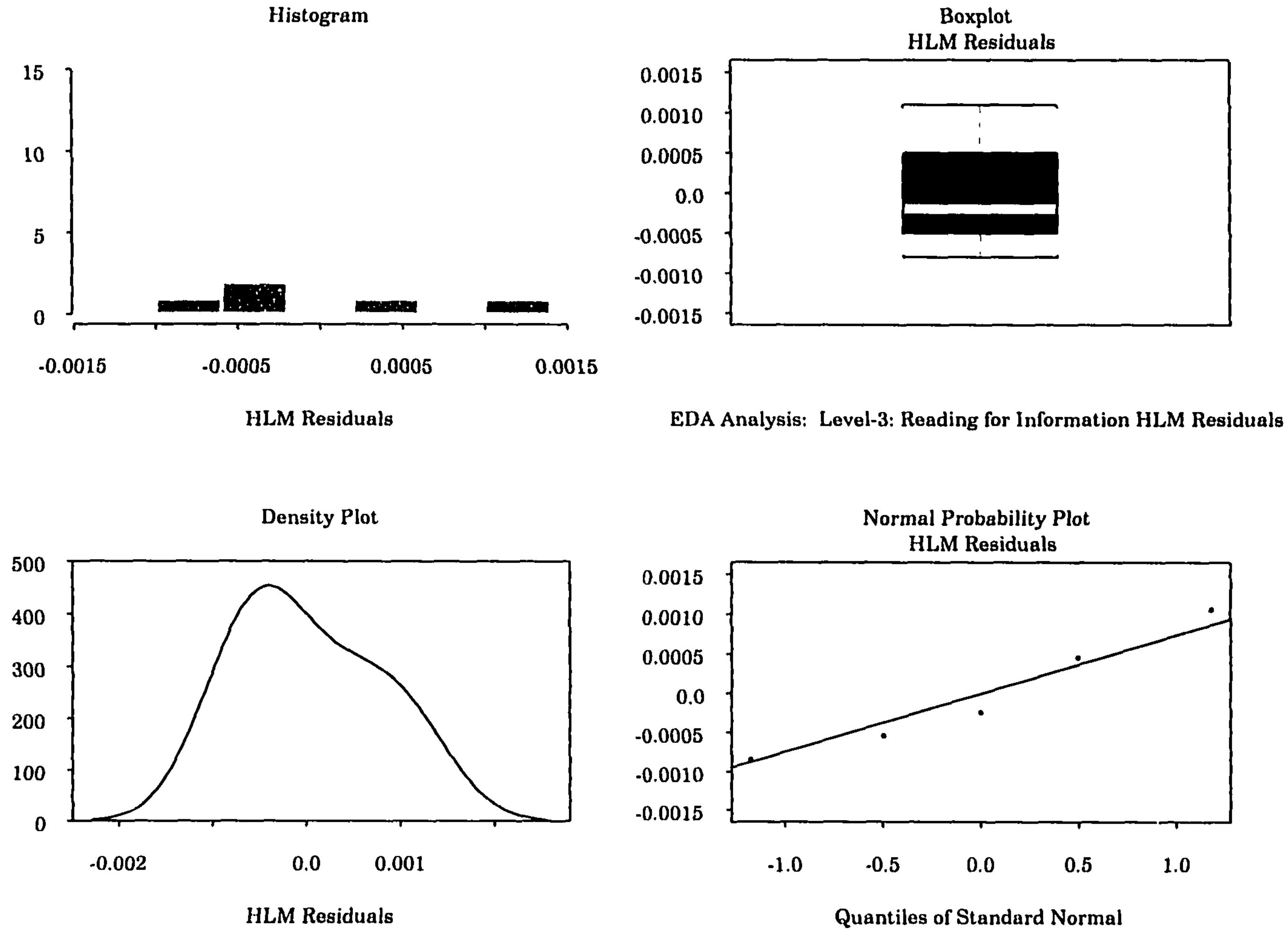


Figure 4.37. Exploratory Data Analysis plots of Reading for Information HLM Level 3 Residuals

Testing the Hypotheses

It is useful at this point to recall the original motive behind the study. The Iowa Department of Education wanted to evaluate the effectiveness of the applied academics component of Iowa's Tech Prep effort. The Request for Proposals focused on two specific questions:

1. Is there a difference in student academic achievement for students who have completed applied academic courses in contrast to those who have completed comparable traditional academic courses?
2. Are the employability skills of students improved by their completion of applied academic courses in contrast to the employability skills of students who have completed comparable traditional academic courses?

Question 1: Academic Achievement

To quantitatively estimate the difference in academic performance between students who have completed applied courses and students who have completed comparable traditional courses (that is, the first research issue), a series of questions and hypotheses were devised. The result of each test is included after each question and statement of the Null Hypothesis. In all paired sample cases, the course means of students enrolled in traditional courses were subtracted from the course means of students in comparable applied courses. A negative difference means traditional students had higher mean scores.

1.1 Is there a statistically significant difference between the paired course mean high school GPAs of students who have completed applied courses versus those who have completed comparable traditional courses?

Null hypothesis: There is no statistically significant difference between the paired course mean high school GPAs of the two groups.

Hypothesis test results: Table 4.18 presents the paired sample data that were used in this test and the results of the test. The Wilcoxon signed-rank test was used because the data showed evidence of a non-normal distribution. The Null Hypothesis was rejected: There is a statistically significant difference (*p value* = 0) between the paired course mean high school GPAs of the two groups. Students enrolled in traditional courses had higher GPAs on average than those enrolled in applied courses. For an excellent summary of the signed-rank test, see Snedecor and Cochran (1989, pp. 140-142).

1.2 Is there a statistically significant difference between the paired course mean composite ITED scores of students who have completed applied courses versus those who have completed comparable traditional courses?

Null hypothesis: There is no statistically significant difference between the paired course mean composite ITED scores of the two groups.

Hypothesis test results: Table 4.19 presents the paired sample data that was used in this test and the results of the test. The Wilcoxon signed-rank test was used because the data showed evidence of a non-normal distribution. The Null Hypothesis was rejected: There is a statistically significant difference (*p value* = 0) between the paired course mean high school ITED scores of the two groups. Students enrolled in traditional courses had higher ITED scores on average than those in applied courses.

Question 2: Impact of Curricula Type

To quantitatively estimate the relationships between curricula and student's employability skills as per the second research question, the independent, dependent, and classificatory variables had to be precisely defined and the above questions rephrased in such a manner as to permit statistical inferences consistent with the assumptions and limitations of the study. The "treatments" under investigation in this study were applied curricula and traditional curricula. The dependent variables used as measures of employability skills included:

- Work Keys Applied Mathematics test score,
- Work Keys Applied Technology test score, and
- Work Keys Reading for Information test score.

The classificatory variables that were ultimately used in the investigation included variables at three levels; student, class, and school. The student level variables included grade level, gender, grade point average, and Iowa Tests of Educational Development composite percentile score (relative to other Iowa students). An additional student-level variable was developed by combining and averaging the GPA and ITED scores. At the class-level, two dichotomous variables were included, type of curricula (applied or traditional), and a variable indicating if the course in which the student was enrolled was relevant to the Work Keys test taken. The school variable was simply an identification number used to account for random variation at that level.

The questions and hypotheses devised to enable the use of statistical inference techniques for the second research question included:

1.1 Is there a statistically significant difference in the raw class mean Work Keys scores between students who have completed applied courses and students who have completed traditional courses?

Null hypothesis: There is no statistically significant difference in the raw class mean Work Keys scores between groups of students who have completed applied courses and students who have completed traditional courses.

Hypothesis test results: Table 4.20 presents the Applied Mathematics paired sample data that was used in this test and the results of the test. The Student's t-test was used since the data appeared to come from a normal distribution. The Null Hypothesis was rejected: There is a statistically significant difference (*p value* = 0.0002) between the paired course mean high school ITED scores of the two groups. Students enrolled in traditional classes scored on average higher than those in applied classes. The 95% confidence interval for the true mean of differences is (-1.096 to -0.460). Table 4.21 presents the Applied Technology paired sample data that was used in this test and the results of the test. The Wilcoxon signed-rank test was used because the data showed evidence of a non-normal distribution. The Null Hypothesis was rejected at a 5% level of significance: There is a statistically significant difference (*p value* = 0.0156) between the paired course mean Applied Technology Work Keys scores of the two groups. Students enrolled in traditional classes scored on average higher than those in applied classes.

Tables 4.22 and 4.23 present the Reading for Information paired sample data that was used in the tests and the results of the tests. Table 4.22 shows the results of the Wilcoxon signed-rank test applied to the original data set which contained an outlier (Traditional Students ID number 4). The Null Hypothesis was not rejected at the 5% level of significance for this data set (*p value* = 0.297). Because of the unusual nature of this outlier and the relatively small sample size comprising this mean, the data were reanalyzed with data pair ID number 4 deleted. Both the Wilcoxon signed-rank test and the Student's t-test were used to evaluate the Null Hypothesis on the new data set. In both these cases, the Null Hypothesis was rejected: The Wilcoxon test indicated a statistically significant difference with *p value* = 0.031; while the Student's t-test yielded a statistically significant *p value* = 0.036 between pairs. With the outlier deleted, results indicated that students enrolled in traditional classes scored on average higher than those in applied classes.

1.2 Are there statistically significant correlations between Work Keys raw class mean scores and the following concomitant student variables: (a) grade level, (b) grade point average (GPA), (c) composite score on the Iowa Tests of Educational Development (ITED), or (d) student gender?

Null hypothesis: No statistically significant correlations exist between Work Keys raw class mean scores and the concomitant variables.

Hypothesis test results: Tables 4.10 through 4.12 present correlation matrices with the 2-tailed Levels of Significance for each of the three Work Keys data sets. The Null Hypothesis was rejected with respect to all three Work Keys tests and all four concomitant variables. In fact, in only one case was the 2-tailed Level of Significance above 0.000. The simple correlation of Applied Math scores with gender yielded a *p value* = 0.004.

2.3 Is there a difference in the adjusted class mean Work Keys scores between students who have completed applied courses and students who have completed traditional courses? The word "adjusted" indicates that the raw Work Keys scores have had the mean effects of any known significant concomitant variables removed.

Null hypothesis: There is no statistically significant difference in the adjusted class mean Work Keys scores between groups of students who have completed applied courses and students who have completed traditional courses.

Hypothesis test results: Hierarchical Linear Models were used to investigate the effect of curricula type on Work Keys scores. Tables 4.31, 4.36, and 4.38 present the results from the analysis of the Applied Math, Applied Technology, and Reading for Information test, respectively. In each of these models a dichotomous variable associated with curricula type (applied or traditional) was included. This variable was a dummy variable associated with curricula type used in classroom j in school k. Applied courses were coded as 0 and traditional courses were coded as 1. A significant positive result for the coefficient of this dummy variable indicates that students in traditional courses score higher on average than do students enrolled in applied courses, after adjusting for significant concomitant variables. In all three cases, significant positive results were obtained and the Null Hypotheses were rejected after adjusting for significant concomitant variables.

The coefficient for the "Curricula gap" in Applied Math was 0.776 with a level of significance (*p value* = 0.000) exceeding the standard 5% level used to reject the Null Hypothesis.

The coefficient for the "Curricula gap" in Applied Technology was 1.035 with a level of significance (*p value* = 0.000) again exceeding the standard 5% level. Of course, the calculated level of significance could be called into question since the data showed evidence of non-normality; however, regardless of the level of significance, the coefficient is positive.

The coefficient for the "Curricula gap" in Reading for Information was 0.826 with a level of significance (*p value* = 0.002); again far exceeding the standard 5% level used to reject the Null Hypothesis.

Discussion

Statistical techniques used in data analysis are built upon a foundation of assumptions and limitations. In this author's opinion, validation of the analysis techniques used in this investigation was at least as important as the findings themselves. For this reason, significant portions of the dissertation dealt with issues such as the use of covariates, choice of units of analysis, problems with non-normal distributions, and intercorrelated independent variables. The assumptions regarding data distributions were checked and residual analyses were performed to ensure that appropriate analysis methods would be chosen. Two methods of analysis were used to investigate the second research question (one method employed covariates, the other did not) since opinions differ as to appropriate use of covariates.

Although a great deal of care went into validation of the analysis methods, one should certainly not take these findings as evidence of a superiority of traditional teaching methods over applied academics, in spite of the statistically significant positive "Curricula gap" coefficients. These were not equivalent groups being compared under true experimental conditions; nor can one discount the possibility of omitted intercorrelated independent variables in the regression equations. An example presented in Neter et al. (1990) helps to clarify this point.

> The fact that intercorrelated independent variables that are omitted from the regression model can influence the regression coefficients in the regression model is illustrated by an analyst who was perplexed about the sign of a regression coefficient in the fitted regression model. He had found in a regression of territory company sales on territory population size, per capita income, and some other independent variables that the confidence interval for the regression coefficient for population size indicated it is negative. The analyst should have considered some of the omitted independent variables in search of an explanation. A consultant noted that the analyst did not include the major competitor's market penetration in the model. Since the competitor was most active and effective in territories with large populations, and thereby kept company sales down in these territories, the result of the omission of this independent variable from the model was a negative coefficient for the population size variable. (p. 301)

As mentioned in Chapter 2: Failure to control certain variables may result in attributing differences in employability skills to instructional method, when in fact one or more of the covariates are correlated with the variables not included in the equation but which are nevertheless related to the dependent variable. A list of such variables might include the socioeconomic status of the student's family, student work load, teacher-to-pupil ratio, availability of course materials, etc. The results presented in this investigation are important in that they examine the impact of several independent variables on the Work Keys test scores, but they should not be generalized as conclusive evidence regarding the overall effectiveness of applied academics. Wang and Owens (1994) outlined a series of criteria in Chapter 2, other than students' test performance, by which applied academic curricula could be evaluated.

A comparison of this investigation relative to others presented in Chapter 2 reveals inconsistent results. Wang and Owens (1995) reported similar academic achievement levels between their Applied Mathematics and comparison group students; whereas in this study, traditional math students had higher levels of academic achievement than students enrolled in Applied Math I and II. Wang and Owens also reported that applied math students scored significantly higher than their peers in traditional math classes, and that Principles of Technology (PT) students performed as well as their traditional counterparts when GPA and grades in math and science were held constant. Neither of the applied groups, PT or math, outscored on average their traditional counterparts in this study. Math and science grades were not included as control variables in this study, as they were in the 1995 Wang and Owens paper. One might reasonably expect these grades to be correlated with overall GPA, but no discussion was provided in their paper as to how they chose control variables. Studies by Dugger & Johnson (1992) and Dugger & Meier (1994) reported greater gains by PT students than their traditional counterparts on Principles of Technology achievement instruments. Although current results obtained by analyzing the Work Keys Applied Technology data showed a significant positive coefficient for students enrolled in traditional courses, one should not forget that the data contained outliers and showed evidence of non-normality, making inferences suspect. The high rate of zero scores by all students, as seen in Table 4.9, was also cause for concern. These facts, coupled with additional information supplied by ACT in their Validity Supplement (1996, pp. 13-14) gives one reason to question the original assumption of a reliable and valid relationship between scores on the Applied Technology test and content-specific employability skills.

CHAPTER 5. CONCLUSIONS

Summary

This dissertation investigates the impact of applied academics on employability skills. In August of 1995, the Iowa Department of Education sent out a Request for Proposals (RFP) for an evaluative study of applied academics in Iowa. This was to be a 10 month project with a start date of October 1, 1995. A proposal submitted by a team from Iowa State University was accepted.

Restatement of the Problem

Employers in Iowa and elsewhere perceive a gap between the level of employability skills of students leaving high school and the level needed to obtain and keep a job in most organizations. Many of Iowa's schools are implementing applied academics courses in an attempt to close this gap. While a considerable amount of anecdotal information exists regarding the effectiveness of applied academics, the impacts of these curricula are not being systematically evaluated to determine if they have an effect on student employability skills.

Restatement of the Purpose of the Study

This study was designed with two goals in mind:

1. Compare the academic achievement, on selected variables, of a sample of Iowa high school students enrolled in applied academics courses against those enrolled in traditional courses.

2. Compare the level of selected employability skills--Applied Math, Applied Technology, and Reading for Information--of a sample of Iowa high school students enrolled in applied academics courses against those enrolled in traditional courses.

Organization of the Study

Chapter 1. Introduction, contains a brief overview of circumstances and driving forces having an impact on the field of education; eventually leading up to current efforts in applied academics. The founding principles of education are briefly reviewed, followed by examples of beliefs, forces, and circumstances that may have guided the evolution of applied academics. A description of the features one currently expects to find in applied academics curricula and an outline of key elements of the Iowa Department of Education RFP appear just before the statement of the problem. The purpose of the study, the research questions, and the assumptions and delimitations of the study follow the problem statement. The Introduction concludes with a section containing the definitions of terms and an overview of the organization of this study.

Chapter 2, the Literature Review, includes a section that gives a historical perspective on Vocational/Technical education and why an evaluation component is needed. Other sections in this chapter cover applied academics research, and statistical methods used in educational research. A summary of the literature review findings completes this chapter.

Chapter 3, Methodology, covers the research approach and design; the population and selection of the sample; the Work Keys instruments used in the analysis; the data collection and analysis procedures; and finally, the assumptions and limitations of the methodology used. A quasi-experimental method of research was used in this investigation.

Chapter 4, titled Results, describes the sample and variables data used in the study; it also contains sections on the results of exploratory data analysis and data analysis for statistical inference; a section discussing the outcomes of the statistical tests; and finally, a discussion segment.

Summary of Findings

Descriptive

The original sample of 1,321 students resulted in full matrix data for 842 students after eliminating series with missing or obviously erroneous data points. Full matrix data for this investigation included information regarding school, type of course (applied or traditional), course, class within course, student gender, student grade in school, student cumulative grade point average (GPA), Iowa Tests of Educational Development score (Iowa percentile rank), the name of the Work Keys Assessment test, and the Work Keys Assessment test score. Since some students took more than one Work Keys test, a total of 1,265 full data series was available for analysis. The remaining 479 students had missing data in one or more of the full matrix data series variables. It was not always necessary to restrict the investigation to those students included in the full matrix data series. Often the entire student sample providing data on one specific variable was used, even if individual members of the sample were unable to provide data on another variable. For example, one could look at the distribution of all students providing GPAs even if some of them did not take the Iowa Tests of Educational Development (ITED). When used, these data were identified as vector data as opposed to the full matrix data. The most common missing independent data were ITED scores; ITED are not universally required and many students simply do not take them.

Graphs in Chapter 4, or Appendix C, include data on grade, gender, GPA, ITED score, and Work Keys score variables compiled over three levels--student (Level 1), class (Level 2), and school (Level 3). Each of these variables compared "applied" students versus "traditional" students in specific subgroups. The first series of Level 1 histograms included all students enrolled in applied courses versus all students enrolled in traditional courses; followed by histograms of students enrolled in specific applied courses versus their counterparts in specific traditional courses. The second series of Level 1 histograms included data on grade, gender, GPA, ITED score, and Work Keys score variables compiled only for those students who scored below the minimum competency level on each of the three Work Keys tests included in the study. These graphs were generated to allow examination of student characteristics for potential patterns in the important subgroup of students who did not meet minimum levels of "employability skills". The Level 2 graphs follow the same pattern as the first series of Level 1 graphs with the exception that class means were charted rather than individual student results. Boxplots were used to compare school-level data collected on applied versus traditional student groups; that is, Level 3 data contain school means for all students enrolled in applied courses versus school means for all students enrolled in traditional courses. Some of the more important general observations contrasting applied versus traditional students include:

- The applied group had more students from the 11th grade and fewer from the 12th grade than the traditional group; 9th and 10th grades had comparable numbers of students.
- More males than females made up the applied group, while slightly more females than males comprised the traditional group.
- Both GPA and ITED histograms showed traditional students with higher means than applied students.
- Higher raw mean scores were also evident on histograms of the traditional groups versus the applied groups on all three Work Keys tests.
- Students scoring below the minimum level of competency on the Work Keys tests were not restricted to those with below average GPA or ITED scores. On

the GPA histograms in particular, the distribution appeared essentially normal and centered approximately on a mean grade of 2.0 on a 4.0 scale.

Some additional observations specific to the three Work Keys tests are listed below:

- The Applied Technology Work Keys test had by far the greatest percent of students scoring below the minimum competency cutoff of 3. Over 41% of all students who took the test scored less than 3. In contrast only 7.2% of all students taking Reading for Information scored below the cutoff and only 2.5% of all students taking Applied Mathematics scored below the cutoff. The group that did the best on the Applied Technology test, the Physics students, had 14.6% of the students below 3. Only one group, at 11.9%, from all those taking either of the other two Work Keys tests had more than 10% of the students scoring below the minimum cutoff.
- More females than males scored below the minimum competency level on the Applied Technology Work Keys test.
- The biggest grade distribution discrepancy was found in the Principles of Technology (PT) versus Physics courses. All Physics students were in the 12th grade, while the PT students were spread out among 10th through 12 grade.
- Of the 152 female students who took the Applied Technology Work Keys test, only 12 were enrolled in Principles of Technology and only 35 were enrolled in

Physics. Of the 232 male students who took the Applied Technology Work Keys test, 74 were enrolled in Principles of Technology (PT) and 75 in Physics. Note: These are numbers from students with complete data series; however the female to male ratios are similar in the vector data series--13 females to 84 males for PT and 43 females to 80 males for Physics.

- The numbers of males and females taking the Applied Math and Reading for Information tests were roughly the same; while the number of males exceeded females (232 to 152) taking the Applied Technology test.
- More males than females (9 to 3) scored below the minimum competency level on the Applied Mathematics Work Keys test.
- Far more males than females (20 to 1) scored below the minimum competency level on the Reading for Information Work Keys test.

Exploratory

The section on exploratory data analysis addressed whether or not the characteristics of the data met those assumed to be true for specific statistical tests. Recall that the first research question addressed differences in academic achievement by evaluating differences in course mean grade point averages and course mean ITED scores. Exploratory Data Analysis (EDA) plots indicated that use of the nonparametric Wilcoxon signed-rank method was preferable to the Student's t-test.

Another series of EDA plots covered course mean Work Keys test scores. Considering these plots, one could use the Student's t-test for the Applied Math data, but the Reading for Information data clearly required the Wilcoxon method. The choice of analysis method for the Applied Technology data was less clear, however since the lower tail of the distribution for students enrolled in traditional courses shows some departure from normality on both the density and normal probability plots, the Wilcoxon method was recommended.

A series of pairwise scatter plots was presented and used to visually evaluate the bivariate relationship of variables considered for use in Hierarchical Linear Models (HLM). In addition to scatter plots, a series of correlation matrices for each Work Keys test was generated. Since there was significant correlation between the ITED scores and GPA for all three Work Keys tests, the decision was made to combine GPA and ITED scores into a new variable called "ACHIEV".

Inferential

Two components were covered in the Statistical Data Analysis section: the first covered questions regarding academic achievement of students in applied courses versus those in traditional courses; and the second, addressed questions concerning the impact of curricula type (applied versus traditional) on employability skills. Differences in academic achievement were evaluated by the Wilcoxon signed-rank test. The impact of curricula type was analyzed two ways;

first using a paired sample test on ***school x course*** Work Keys test means, and second using Hierarchical Linear Model (HLM) techniques. The results of the paired sample tests were as follows:

- **Question:** Is there a statistically significant difference between the paired course mean high school GPAs of students who have completed applied courses versus those who have completed comparable traditional courses?

  Answer: There was a statistically significant difference (p *value* = 0) between the paired course mean high school GPAs of the two groups. The course mean GPAs for the traditional students were higher on average than the course mean GPAs for applied students.

- **Question:** Is there a statistically significant difference between the paired course mean composite ITED scores of students who have completed applied courses versus those who have completed comparable traditional courses?

  Answer: There was a statistically significant difference (p *value* = 0) between the paired course mean high school ITED scores of the two groups. The course mean ITED scores for the traditional students were higher on average than the course mean ITED scores for applied students.

- **Question:** Is there a statistically significant difference in the raw class mean Work Keys scores between students who have completed applied courses and students who have completed traditional courses?

Answer: There was a statistically significant difference (*p value* = 0.0002) between the paired course mean Applied Math scores of the two groups. The course mean Applied Math scores for the traditional students were higher on average than the course mean Applied Math scores for applied students. There was a statistically significant difference (*p value* = 0.0156) between the paired course mean Applied Technology scores of the two groups. The course mean Applied Technology scores for the traditional students were higher on average than the course mean Applied Technology scores for applied students. There was not a statistically significant difference at the 5% level (*p value* = 0.297) between the paired course mean Reading for Information scores of the two groups containing one outlier in the traditional group. When the data pair containing this outlier was removed and the data set reanalyzed there was a statistically significant difference (*p value* = 0.036) between the paired course mean Reading for Information scores of the two groups. The course mean Reading for Information scores for the traditional students were higher on average than the course mean Reading for Information scores for applied students.

- **Question:** Are there statistically significant correlations between Work Keys raw class mean scores and the following concomitant variables: (a) student grade level, (b) grade point average (GPA), (c) composite score on the Iowa Tests of Educational Development (ITED), or (d) gender?

Answer: There are statistically significant correlations with respect to all three Work Keys tests and all four concomitant variables. In fact, in only one case was the 2-tailed Level of Significance above 0.000. The simple correlation of Applied Math scores with gender yielded a p value = 0.004.

- **Question:** Is there a difference in the adjusted class mean Work Keys scores between students who have completed applied courses and students who have completed traditional courses? The word "adjusted" indicates that the raw Work Keys scores have had the mean effects of any known significant concomitant variables removed.

Answer: Hierarchical Linear Models were used to investigate the effect of curricula type on Work Keys scores. In each of these models a dichotomous variable associated with curricula type (applied or traditional) was included. This variable was a dummy variable associated with curricula type used in classroom j in school k. Applied courses were coded as 0 and traditional courses were coded as 1. A significant positive coefficient for this dummy variable indicated that students in traditional courses scored higher on average than did students enrolled in applied courses, after adjusting for other significant concomitant variables. In all three cases, significant positive coefficients were obtained. The coefficient for the Applied Mathematics assessment test was 0.776 with a *p value* = 0.000 level of significance. The coefficient for the Reading for Information assessment test was 0.826 with a *p* *value* = 0.002 level of significance. The coefficient for the Applied Technology assessment test was 1.035 with a *p value* = 0.000 level of significance. It should be noted, however, that the calculated level of significance for the Applied Technology coefficient could be called into question, since the data showed evidence of non-normality.

Conclusions

First, lest anyone come to the conclusion based on this investigation that applied academics should be discarded in favor of traditional teaching methods: One should certainly not take these findings as evidence of a superiority of traditional teaching methods over applied academics, in spite of the statistically significant positive "Curricula gap" coefficients. These were not equivalent groups being compared under true experimental conditions; nor can one discount the possibility of omitted intercorrelated independent variables in the regression equations. Failure to control certain variables may result in attributing differences in employability skills to instructional method, when in fact one or more of the covariates are correlated with the variables not included in the equation but which are nevertheless related to the dependent variable. A list might include socioeconomic status of the student's family, student work load, teacher-to-pupil ratio, availability of course materials, or a host of other variables. The results presented in this investigation are important in that they examine the impact of several independent variables on the Work Keys test scores, but they should not be generalized as conclusive evidence regarding the overall effectiveness of applied academics. Dugger and Johnson (1992) noted during their comparison of physics and Principles of Technology programs that "One must exercise caution in drawing inferences regarding the two programs since physics also is responsible for covering higher level concepts that are not considered basic and may be considered non-intuitive" (p. 25). Wang and Owens (1994), when reporting characteristics of applied curricula, stated that they, "are not meant to replace 'traditional' academic courses for the top 25 percent of the student population" (as cited in Limback & Rosa, 1996, p. 151). Both these statements reinforce the fact that non-equivalent groups are being compared.

After (hopefully) having made point that one cannot conclude that traditional teaching methods are superior to applied academics, one may well ask, what conclusions can be drawn about the sample in this study.

- The two groups, students enrolled in applied academics course and those enrolled in comparable traditional courses, do differ in levels of academic achievement as measured by GPA and ITED scores, as well as by raw score performance on two of the three Work Keys tests. The one exception to difference in levels of performance on a Work Keys test was driven by one outlier whose mean value was made up of a limited number of students with extreme test scores. If one excludes that pair of points (which seems reasonable), then the differences in performance for all three Work Keys tests would be significant. In all cases exhibiting statistically significant results, traditional students' performances exceeded on average performances of the applied students.

- The differences between student performance on the Applied Technology Work Keys test and the other two Work Keys tests are large enough to warrant investigation. The number of students scoring less than the minimum competency cutoff score on the Applied Technology test was far too high to ignore.
- A gender gap exists in the technology courses; that is, physics and Principles of Technology. Far more males than females are enrolled in these courses.
- The use of a simple "relevant course" variable was inadequate for the Applied Math and Reading for Information data. The coefficient was negative for the Applied Math data and insignificant for the Reading for Information data. It was positive for the Applied Technology data, but expectations were that it would be positive for all three sets of data. One possible explanation for this is that most students are required to take some math and English courses; while Principles of Technology and physics are electives. A negative coefficient could occur for the Applied Math data, for example, if students chosen for the study because of their enrollment in an English course, were simultaneously enrolled in an advanced "relevant" math course.

- Hierarchical Linear Modeling appears to offer great promise as an analytical technique for the kind of nested data found in this investigation.
- The choice of independent variables in regression equations is a matter of no small importance. Analysis of the Applied Technology data highlighted a classic problem in regression--utilizing intercorrelated independent variables.
- A researcher must be aware of the assumptions upon which certain statistical techniques are based and check to see if those assumptions are indeed realistic.
- Much of the value of the information in this study may lie in its use as baseline data for subsequent studies.

Implications for Iowa

The question as to the impact of applied academics on employability skills is a complex issue that was only partially answered by this investigation. Research is an iterative process, with initial studies often raising more questions than they answer. The investigation provided insight into certain group demographics and the relationships of certain variables to Work Keys scores; however it also raised questions as to how one should assess the effectiveness of applied academics curricula. Given the findings of this investigation, one would argue that the effectiveness of applied academics cannot be determined solely from simple test scores. Although traditional students did better on average on the Work Keys tests than did applied students, even after controlling for a number of concomitant variables, test scores are not the only indicators of employability skills. Proponents of applied academics have suggested additional measures of effectiveness. Several from a list by Hull and colleagues are included below:

- Students are able to transfer knowledge from academic content to vocational applications and from school to the workplace.
- Students are not afraid to take academic subjects such as mathematics and science.
- Students display more interest, motivation, and understanding of the value of the subject and of school in general than they did in classes taught by traditional methods. ...
- The student population that has traditionally done poorly in academic subjects displays improved performance. (as cited in Wang & Owens, 1994, p. 8)

So what then are the implications of this investigation for Iowa? There is certainly nothing in this investigation to indicate that applied academics are not effective when measured against objectives such as those listed above. Indeed, based on lessons learned during this investigation, one may conclude that to fully study the effectiveness of applied academics the following must occur:

1. Include other measures of employability skills. Work Keys assessment tests may serve as one measure of these skills, however they have yet to gain widespread acceptance as true measures of employability skills. In addition, multiple sites expressed concern regarding cost of the test instruments.
2. Broaden the investigation to include additional measures of effectiveness, such as those suggested by Hull and colleagues.

3. Improve the data collection system statewide. The method used to collect data for this investigation was time-consuming and yielded less-than-optimum results. The original team struggled to resolve issues associated with incomplete data throughout the project. This dissertation contains data that the original report to the Iowa Department of Education does not, simply because the data were unavailable at the time the report was submitted.
4. Monitor growth of students' employability skills over time. Data should be collected at periodic intervals for analysis and should include measures of performance in both school and workplace. Data collected and analyzed during this investigation are the start of a reasonable set of baseline numbers, but repeated observations are needed. Applied academics curricula are relatively new additions to many school systems and a "learning curve" exists with respect to their implementation. Tracking changes over time is of particular importance if the measures of effectiveness suggested by Hull and colleagues are to be examined. Care must also be taken in the choice of when data are to be collected. One data collection site reported that students, particularly seniors, were less apt to put forth their best efforts when tests were administered near the end of the school year.
5. Investigate other independent variables that may account for the significant unexplained variability related to Work Keys test scores. Besides some potential variables already mentioned, such as socioeconomic status, existing

variables might be further refined; for example, instead of using composite ITED and GPA for all Work Keys tests, ITED math subscores and GPA for math courses only might be used as independent variables for the Applied Math test.

6. Include a broader selection of high schools in any future studies. The results of this investigation should not be generalized to all Iowa high schools due to limited sample size and lack of an adequate cross-section.
7. Finally, take care in the choice of statistical methods to be used for analyses. This dissertation has attempted to point out some of the issues and pitfalls associated with the selection and use of specific statistical models.

The above discussion focuses on the ability of researchers to evaluate, or check, the effectiveness of applied academics curricula with respect to employability skills. This is essentially the third component of the classic Shewhart cycle discussed by Deming (1986, p. 88) and, in adapted form, shown in Figure 5.1. Development and implementation of the applied academics curricula are components one and two, respectively. The remaining, fourth, component of the cycle refers to the actions one takes based upon the results of the evaluation. Clearly, additional work is needed to develop and implement an evaluation process with which everyone is comfortable; however , with the initial evaluation efforts completed, the burden of responsibility shifts to members of the community to act upon the results of the evaluation process. Educators and employers together have a responsibility to act. It was reported in Chapter 1, the

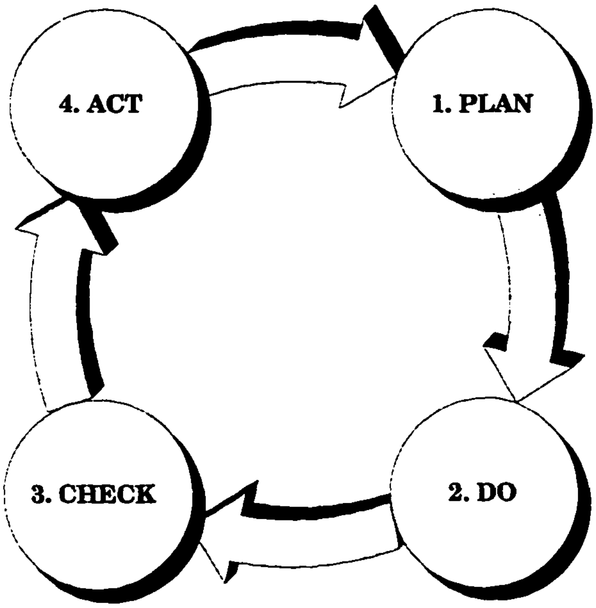


Figure 5.1. Shewhart cycle.

Introduction, that employers should:

> ... tell educators clearly what you need and work with them to accomplish it. You know that students have to believe that you care about what they learn. Employers who value performance in high school when they make their hiring decisions provide students with the right signal: learning and earning are related activities.
>
> Finally, ... confirm that the SCANS skills accurately reflect your local workforce requirements. Having confirmed these skills, make sure your local school board never loses sight of them in instructional planning. (Secretary's Commission on Achieving Necessary Skills, 1991, p. viii)

These same recommendations hold true regarding applied academics curricula.

Recommendations for Future Research

The following recommendations are offered with the objective of better assessing the effectiveness of applied academics; they are therefore tied closely to the list outlined in the previous section:

1. Develop and complete a multi-year study to evaluate students' growth in employability skills over a 5-year time span (8th grade through 12th grade). One possible design would call for each student to be measured using Work Keys assessment tests at the mid-point of each school year over the time span of the study. The number of students tested in each class should exceed 10 to be certain that at least 10 good data sets are available for analysis. This level of within-class participation is needed to meet the requirements of some of the HLM techniques used in the analysis process, such as testing for homogeneity of variance. Enrollment in relevant classes, grade level, gender, type of class, and some type of academic achievement variables, along with test results, would be recorded throughout the time covered by the study. Other potentially important variables, such as socioeconomic status of the student, could also be included in this study. This investigation would allow an estimation of the correlation between the initial status of the students and their rate of change regarding employability skills. It would also allow a comparison between the rate of change of students enrolled in applied academics courses with those enrolled in comparable traditional courses. For those students entering the workforce directly after graduation from high school, additional data should be collected, including an evaluation by employers of the new employees' entry-level workplace skills. Tracking changes over time is of particular importance if certain additional measures of effectiveness, such as those suggested by Hull, are to be examined.

2. Investigate other measures of employability skills. Work Keys assessment tests may serve as one measure of these skills, however they have yet to gain widespread acceptance as true measures of employability skills. In addition, multiple sites expressed concern regarding cost of the test instruments.
3. Consider measures of the effectiveness of applied academics other than simple test scores, such as those suggested by Hull and colleagues.
4. Investigate other independent variables that may account for the significant unexplained variability related to Work Keys test scores. Besides some potential variables already mentioned, such as socioeconomic status, existing variables might be further refined; for example, instead of using composite ITED and GPA for all Work Keys tests, ITED math subscores and GPA for math courses only might be used.
5. Replicate the study using stratified random sampling, if necessary, to include additional schools districts in different areas. This study was essentially restricted to a band that ran from the northwest corner of the state to the southeast corner of the state. Additional information regarding schools in the northeast and southwest segments of the state would be helpful. Information regarding performance of large metropolitan schools would also be of benefit.
6. Conduct follow-up studies at the school originally involved with this investigation. They have provided valuable baseline data and continued studies would provide them with feedback on program progress. If resources

for follow-up studies are fixed at the same level as for this investigation, the number of students within a class to be tested should be reduced; while the number of classes and schools should be increased. The "experimental unit" here is the class, not the individual within the class.

7. Initiate an investigation into the cause of the high number of students scoring below the minimum competency level on the Applied Technology test. If the Work Keys Applied Technology test truly measures skills needed in the workplace, then this offers a tremendous opportunity for feedback to the schools regarding needed change. If there is a problem with the test, then the sooner that feedback is provided to the developers, the more quickly it can be corrected.

APPENDIX A: WORK KEYS INSTRUCTIONS

High School: ______________________

IOWA APPLIED ACADEMICS EVALUATION STUDY: INSTRUCTIONS FOR HANDLING, ADMINISTERING, AND RETURNING WORK KEYS MATERIALS

Thank you for your participation in this important study. The instructions that follow have been tailored, with permission, for your use from the Work Keys Administrator's Manual (The American College Testing Program, 1995). We trust that these instructions will be straightforward and simplify the Work Keys administration and handling process at your school. If you have questions about anything herein or you have a case not covered by the instructions, please **phone (515/294-7009) or FAX a message (515/294-9284) to Mari Kemis or Mandi Lively at Iowa State University.**

CHECK-IN AND SECURITY OF ASSESSMENT MATERIALS

Open the cartons in your shipment of assessment materials and check the contents against the list below:

________ copies of the "Applied Mathematics" test

________ Math Formula Sheets, for use with "Applied Mathematics" test

________ copies of the "Applied Technology" test

________ copies of the "Reading for Information" test

________ sets of student questions for Local Items block of answer document

________ sets of Work Keys Answer documents

________ copies of "Work Keys Job Header or Group/Class Header," one for each course being tested by a Work Keys test, with the course and school name recorded in Block "A," course name coded in Block "B" with a #2 pencil, and number of students to be tested in this course coded in Block "C"

__One___ copy of "Work Keys Company Header or Building Header" with the school contact person's name, high school name and other information printed in Block "A;" and school name, AEA number, "building" and number of answer folders pre-coded in Blocks "B-E" with a #2 pencil

________ pre-paid and pre-addressed United Parcel label(s) for returning the answer documents, test booklets and all other assessment materials in the same carton(s) sent to you.

If there is any discrepancy in quantities or any evidence of tampering, report it to Mari or Mandi at the phone/FAX number above. Exercise vigilance concerning the security of all assessment materials before, during, and after administering the assessment. **All assessment booklets must be accounted for before and after testing and returned to RISE at Iowa State University.** Any security breach (theft or loss) should be reported immediately at the telephone number identified in the opening paragraph.

After you have verified receipt of all materials, reseal the cartons and lock in a secure place to which only you (or you and a few specifically authorized Persons) have access. Protect the materials from damage, from possible theft or loss, and from any conditions that could allow prior knowledge of the assessments.

If more than one testing room is to be used, count out the appropriate quantities of materials before testing and record the number of assessment booklets assigned to each room.

Save the carton(s) in which the materials were shipped. It will be required for returning all materials at the completion of the testing, and have postage-paid labels attached.

Give assessment booklets to the room supervisors personally. Never leave booklets in an unattended room.

ARRANGEMENTS FOR TESTING

Day and Time of Testing

Objective test assessment testing times range from **40 to 45 minutes**, depending on the assessment. Additional time of 25-40 minutes is required for completing demographic information and the "Local Items" block on the answer documents, and could be completed at the end of a previous class period.

Our understanding is that you will be administering each test separately during the normal meeting time for that course, or to all students in the applied academics course and the "comparable course" together in an assembly setting. The same verbal instructions to students will apply for either method. Be sure to allow examinees the appropriate amount of time to complete each assessment.

Make-up Testing

Make-up testing can be administered if a small number of examinees are unable to test during the designated testing schedule. Reasons may include schedule conflict or illness.

Testing Staff

If the Work Keys assessments will be administered in more than one room, you will need a room supervisor in each room to read the directions to the examinees and to monitor examinee activities. Depending on the number of examinees in each room, you may wish to use proctors to assist the room supervisors in distributing and collecting assessment materials and monitoring the test room. It is recommended that you use a proctor when more than 30 examinees are in the group, and when examinees who are potentially disruptive or have special needs are present. **Be sure that all personnel assisting you are familiar with the contents of these instruction sheets.**

Test Rooms

Select test rooms that offer adequate writing surfaces, uncrowded seating, good lighting, comfortable temperatures, a quiet atmosphere, and freedom from distraction.

Seating Arrangements

Whenever possible, seat examinees at separate desks in a block so that all rows (side-to-side) and columns (front-to-back) have about the same number of examinees. This arrangement simplifies the distribution, 2 collection, and verification of assessment materials. Make sure that all examinees face the same direction. Arrange seating to prevent examinees from communicating or looking at one another's answer document. Always assign examinees to their seats; do not allow them to choose their own. There should be at least **three feet** of space between examinees. If elevated seating is used (e.g., in a tiered auditorium) also provide a minimum distance of five feet from front to back. If the desks or chairs are stationary, seat examinees in every other column and make sure that examinees are seated directly behind one another. If the seats are movable, you may use them all, provided they are three feet apart and in straight columns and rows. Be sure the aisles between rows or columns of seats are wide enough for testing personnel to circulate during the examination without disturbing examinees.

Writing Surfaces

Writing surfaces must be large enough to accommodate the assessment booklet and answer document side by side. Examinees should not be distracted by inadequate writing surfaces. Lap boards may not be used.

Bulletin Boards

Check each testing room to make sure that maps, periodic tables, posters, charts, and bulletin board materials related to subjects of the assessments are removed or covered.

Materials Supplied by the Test Site or Examinee

- A reliable stopwatch or interval timer for each test room.
- A supply of soft-lead (No. 2) pencils with erasers to lend to examinees who do not bring pencils. Instruct examinees prior to testing to bring two (No. 2) lead pencils with them on test day(s).
- A pencil sharpener.
- Calculators to lend to examinees who do not bring one for the Applied Mathematics assessment.
- Social Security numbers. Instruct examinees prior to testing to bring their social security numbers with them on test day.

Defective Assessment Booklet or Answer Document

If a defective assessment booklet is discovered, immediately replace the booklet with another. Write the nature of the defect on the cover of the booklet and note the defect on the Irregularity

Report. Attach the Testing Irregularity Report to the defective booklet and return it with the other test booklets (see "Return of Answer Documents and Assessment Materials" section).

If a defective answer sheet is discovered, immediately replace it with a new one. Have the examinee transfer all previously written information to the new answer document **after** the timed portion of testing is completed. Complete a Testing Irregularity Report, attach the defective answer document, and return it as described in the "Return of Answer Documents and Assessment Materials" section.

Voiding Single Assessments

If one or more of the individual assessments gridded on an answer document should not be scored, write VOID THIS TEST, in red, across the specific assessment block (s). Make an entry on the Testing Irregularity Report and attach the answer document with a paper clip. Return the Testing Irregularity Report and attached answer documents with the other scorable answer documents as described in the "Return of Answer Documents and Assessment Materials" section.

Examinee Who Becomes Ill

If an examinee becomes ill, dismiss the examinee from the test room and mark the assessment section VOID (see the "Voiding Single Assessments" subsection below).

Testing Irregularity Report

The Testing Irregularity Report is intended for use as a record of any test administration irregularities that may affect examinee scores or the interpretation of the Work Key results, or that result in voiding one or more assessments. This form is presented on page 8, and as many copies should be made as needed. Room supervisors and test administrators should use the form to report any of the irregularities listed below:

- An examinee becomes ill and discontinues testing.
- An examinee is giving or receiving assistance, or is filling in ovals after final time is called.
- The assessment is mistimed.
- An examinee is using an unauthorized testing aid.
- A disturbance or distraction occurs that could affect one or more examinee scores.
- An examinee questions an item or scenario.
- An examinee has a defective assessment booklet or answer document.

PLANNING YOUR TEST ADMINISTRATION SCHEDULE

On the Planning Chart below, identify the assessments you plan to administer and indicate the order in which you plan to administer them.

PLANNING CHART

Name of Assessment	Administration Time (in minutes)	Number of Items	Administration Order of Assessments	Administration Directions on Pages
Completing Answer Document Demographics and Local Items	25-40	N/A	1	11-14
Applied Mathematics	40	30		15-19
Applied Technology	45	32		20-23
Reading for Information	40	30		24-27

Time Available

Make a copy of a Session Chart for each assessment you plan to administer (two charts are provided on Page 9 for your convenience). Number the sessions to match the order in which you will administer the assessments. Remember to include a session for completing the answer document demographics and Local Items section (handing out the sets of Local Items with the answer sheet); for example, this could be scheduled at the end of a previous class period for the course. If your testing time is constrained (e.g., by class periods), enter the total time available in the last line for each block.

Actual Testing Time

On each Session Chart, enter the name of the assessment to be administered and the appropriate administration time (from the Planning Chart above).

Breaks

If you are administering several sessions "back-to-back" in a continuous schedule, you should provide breaks between sessions. If you are giving only two assessments, 5-10 minutes is sufficient. Determine your break times and add the minutes to the session.

You may use the five minutes to collect and verify materials as a break if examinees simply stand and stretch in place. Be sure examinees turn their assessment materials face down. However, if examinees are moving around the room or going outside of the room for a break, for security reasons you must collect all of the assessment materials before allowing examinees to leave their seats.

Total Time

Calculate the total time required in each block and compare it to the total time available. Make any adjustments to ensure that you have a realistic schedule. If possible, build a few extra minutes into each session for examinee questions or other unanticipated events.

TEST ADMINISTRATION PROCEDURES

Administrator's Instructions

Your examinees will use the ACT Work Keys answer documents. It is **very important** that all testing personnel be familiar with the instructions on completing the personal demographic information, **Test Form Numbers**, and **Booklet Numbers** on the answer documents. The Test Form Numbers indicate which answer key ACT will use in scoring the assessment. Therefore, **if a Test Form Number is not entered correctly, ACT cannot score the answer document correctly**.

Once testing has begun, do not admit examinees who arrive late unless you can provide them with appropriate directions and the full testing time without disrupting the other examinees.

Instructions Before Test Day

A few days prior to the test day(s), announce to examinees that they must bring the following:

- Two soft-lead (No. 2) pencils with erasers
- A watch if they wish to pace themselves (They must not set the alarm on the watch during the assessment.)
- Social security number
- A calculator for the Applied Mathematics Assessment

Avoiding Common Errors in Completing Answer Documents

Room supervisors should be alert to the types of errors examinees commonly make when completing their answer documents. When reading the instructions, test personnel should emphasize the correct procedures to avoid these errors, walking around the testing room to observe examinees as they complete these steps.

When completing the demographic information:

- grid only one oval per column.
- start with the first box and first column of ovals in the block.
- grid name and address in addition to writing them in the spaces.

When completing each assessment section:

- grid the appropriate test form code in addition to writing the code and name in the spaces provided.
- grid administration codes in addition to writing them in the designated spaces.
- emphasize that examinees must mark their responses on the answer document, not the assessment booklet. No additional time will be allowed for transferring answers marked in assessment booklets unless an assessment accommodation is used.

RETURN OF ANSWER DOCUMENTS AND ASSESSMENT MATERIALS

Room Supervisor Tasks

To ensure that each examinee's assessment results are reported accurately and quickly as possible, at the completion of testing, you should assemble and check the answer documents carefully.

- Make sure that there is an answer document for each examinee who took an assessment.
- Check each examinee's answer document to note the following:
  --Is each examinee's name printed and gridded properly?
  --Is all other required student information (e.g., Social Security number, birth date) complete/accurate?
  --Is the Local Items block on the answer document gridded properly?
  --Has each examinee completed the test booklet number and test form code key for each Work Keys assessment administered?
  --Are answers marked with a soft-lead pencil? If an examinee used a pen or marker, use a soft-lead, No. 2 pencil to grid over the ink marks.
  --Are all marks neat, dark, and gridded properly, and have all stray marks or doodles been erased?
- Keep other used and unused assessment materials separate, including Work Keys assessment booklets, Applied Mathematics formula sheets and unused answer documents, and return all materials to your test coordinator for return to the state-wide study coordinator Iowa State University.
- Use a copy of the Testing Irregularity Report form on page 8 to describe any irregularities which could affect the examinee's scores. Place it on top of the answer documents and return to test coordinator.

Test Coordinator Tasks

- Collect all answer documents for a testing site.
- Ensure that all answer documents have been received from each testing site.
- Separate your answer documents according to course groups and place the appropriate "Job Group/Class Header" **on top of each set of answer documents**, with the "Company Header/Building Header" (high school name gridded on it) **on top of the total combined group of answer documents.**
- Place the answer documents into the envelope(s) provided (making certain all members of a group are in the same envelope). Number the envelopes sequentially (i.e., 1 of 5, 2 of 5, 3 of 5, etc.), and on each envelope list the groups for which answer documents and a header are contained in that envelope.
- Place all forms completed by the Test Administrator(s) on top of the materials in the first envelope. These may include any Testing Irregularity Reports and a Report of Administration Accommodations.
- Ensure all non-scorable test material has been collected from every testing site.
- Place <u>all</u> of the non-scorable materials (test booklets and other assessment materials that will not be scanned) into the bottom of the cartons(s) used to ship the test materials to your high school, and place the envelopes with answer documents at the top of the carton. If more than one carton was used to mail the materials to you, the cartons will be numbered consecutively with magic marker; place the envelopes containing your answer documents at the top of Carton No. 1.
- Place the pre-paid and pre-addressed United Parcel Service (UPS) label(s) onto the carton(s) for returning the answer documents/test booklets/other assessment materials, and ship by UPS to RISE at Iowa State University.

APPENDIX B: GENERAL HIERARCHICAL LINEAR MODELS

This appendix contains examples of possible two-level conditional models. These do not necessarily cover the model that will ultimately survive when the questions of which predictor variables to include in the model and how their coefficients should be specified (fixed, random, nonrandomly varying) are answered. One would arrive at a two-level model if the Level-3 (school) variation was small enough to "ignore". For purposes of discussion regarding methodology, a two-level model will be presented below; although the Level-3 (school) variation will be checked. Examples of possible fixed, random, and nonrandomly varying coefficients are shown at Level-2.

General Level-1 Model: Within each classroom, we model student employability skills (that is, Work Keys assessment test scores) as a function of the student-level predictors; here ITED composite score, gender, and grade level, plus a random student-level error:

$$Y_{(ij)} = \pi_{0(j)} + \pi_{1(j)}a_{1(ij)} + \pi_{2(j)}a_{2(ij)} + \pi_{3(j)}a_{3(ij)} + e_{(ij)}$$

Where

$Y_{(ij)}$ is the Work Keys test score of student *i* in classroom *j*.

$\pi_{0(j)}$ is the mean Work Keys test score of 9th grade females in classroom *j*.

$\pi_{1(j)}$ is the predicted change to mean Work Keys test score in classroom *j* per unit change in the student's class (or grade) centered composite ITED score.

$a_{1(ij)}$ is the class (or grade) centered composite ITED score of student i in classroom j.

$\pi_{2(j)}$ is the predicted change to mean Work Keys test score in classroom j when the student is a male. This is a "gender-gap" coefficient.

$a_{2(ij)}$ is a dummy variable associated with student gender. The coding is 0 for a female student and 1 for a male student.

$\pi_{3(j)}$ is the predicted change to mean Work Keys test score in classroom j as a result of the student's grade level (9th, 10th, 11th, or 12th grade).

$a_{3(ij)}$ is a dummy variable associated with student grade level. The coding is 0 for a student in 9th grade, 1 for a student in 10th grade, 2 for a student in 11th grade, and 3 for a student in 12th grade.

$e_{(ij)}$ is a Level-1 random effect that represents the deviation of student ij's score from the predicted score. These residual effects are assumed normally distributed with a mean of 0 and a variance of σ^2.

General Level-2 Model: Each of the regression coefficients in the above Level-1 model (including the intercept) can be viewed as either fixed, nonrandomly varying, or random. In addition each Level-1 coefficient may be predicted or modeled by some classroom-level characteristics. This leads to the following general formulation of the model for variation among classrooms.

All fixed:

$$\pi_{0(j)} = \beta_{00}$$

$$\pi_{1(j)} = \beta_{10}$$

$$\pi_{2(j)} = \beta_{20}$$

$$\pi_{3(j)} = \beta_{30}$$

All nonrandomly varying:

$$\pi_{0(j)} = \beta_{00} + \beta_{01}X_{1(j)}$$

$$\pi_{1(j)} = \beta_{10} + \beta_{12}X_{2(j)}$$

$$\pi_{2(j)} = \beta_{20} + \beta_{21}X_{1(j)} + \beta_{22}X_{2(j)}$$

$$\pi_{3(j)} = \beta_{30} + \beta_{31}X_{1(j)} + \beta_{32}X_{2(j)}$$

All random:

$$\pi_{0(j)} = \beta_{00} + \beta_{01}X_{1(j)} + \beta_{02}X_{2(j)} + r_{0(j)}$$

$$\pi_{1(j)} = \beta_{10} + \beta_{11}X_{1(j)} + \beta_{12}X_{2(j)} + r_{1(j)}$$

$$\pi_{2(j)} = \beta_{20} + \beta_{21}X_{1(j)} + \beta_{22}X_{2(j)} + r_{2(j)}$$

$$\pi_{3(j)} = \beta_{30} + \beta_{31}X_{1(j)} + \beta_{32}X_{2(j)} + r_{3(j)}$$

Where

$X_{1(j)}$ is the grand mean-centered composite ITED score of classroom *j*.

$X_{2(j)}$ is a dummy variable associated with curriculum type used in classroom *j*. The coding is 0 for an applied course and 1 for a traditional course.

β_{00} is the class (or grade) mean Work Keys test score of females in applied curricula.

β_{01} is the predicted change to overall class- (or grade-) mean Work Keys test score of females per unit change in the class- (or grade-) grand mean-centered composite ITED score.

β_{02} is the predicted change to overall class- (or grade-) mean Work Keys test score of females in classroom *j* when traditional curricula are used rather than applied curricula. This is a "curricula-gap" coefficient.

$r_{0(j)}$ is a Level-2 random effect that represents the deviation of class (or grade) *j*'s Level-1 coefficient from its predicted value based on the Level-2 model. The random effects in these equations are assumed to be correlated. They are also assumed to be multivariate normally distributed with a mean of 0. The variance of this effect is designated as τ_{00} and has covariances τ_{01} and τ_{02} with the other Level-2 random effects.

β_{10} is the predicted mean change to $\pi_{1(j)}$ (slope relating student ITED to Work Keys score) when the class- (or grade-) mean composite ITED score is the same as the grand-mean composite ITED score and applied curricula are being used. It is a mean since we are using grand mean-centered ITED scores for $X_{1(j)}$. Note that when class- (or grade-) mean ITED score is equal to grand mean ITED score $X_{1(j)} = 0$.

β_{11} is the predicted mean change to $\pi_{1(j)}$ (slope relating student ITED to Work Keys score) per unit deviation in the class- (or grade-) mean composite ITED score from the grand-mean composite ITED score.

$\beta_{12(k)}$. is the predicted mean change to $\pi_{1(j)}$ (slope relating student ITED to Work Keys score) when traditional curricula are used rather than applied curricula.

$r_{1(j)}$ is a Level-2 random effect that represents the deviation of class (or grade) *j*'s slope coefficient (relating student ITED to Work Keys score) from its predicted value based on the Level-2 model. The random effects in these equations are assumed to be correlated. They are also assumed to be multivariate normally distributed with a mean of 0. The variance of this effect is designated as τ_{10} and has covariances τ_{11} and τ_{12} with the other Level-2 random effects.

β_{20} is the predicted mean change to $\pi_{2(j)}$ (slope relating student gender to Work Keys score) when the student is male and two conditions hold: (1) the class- (or grade-) mean composite ITED score is the same as the grand-mean composite ITED score and (2) applied curricula are being used.

β_{21} is the predicted mean change to $\pi_{2(j)}$ (slope relating student gender to Work Keys score) per unit deviation in the class- (or grade-) mean composite ITED score from the grand-mean composite ITED score.

β_{22} is the predicted mean change to $\pi_{2(j)}$ (slope relating student gender to Work Keys score) when traditional curricula are used rather than applied curricula.

$r_{2(j)}$ is a Level-2 random effect that represents the deviation of class (or grade) *j*'s slope coefficient (relating student ITED to Work Keys score) from its predicted value based on the Level-2 model. The random effects in these equations are assumed to be correlated. They are also assumed to be multivariate normally distributed with a mean of 0. The variance of this effect is designated as τ_{20} and has covariances τ_{21} and τ_{22} with the other Level-2 random effects.

β_{30} is the predicted mean change to $\pi_{3(j)}$ (slope relating student grade level to Work Keys score) when the student is male and two conditions hold: (1) the class- (or grade-) mean composite ITED score is the same as the grand-mean composite ITED score and (2) applied curricula are being used.

β_{31} is the predicted mean change to $\pi_{3(j)}$ (slope relating student grade level to Work Keys score) per unit deviation in the class- (or grade-) mean composite ITED score from the grand-mean composite ITED score.

β_{32} is the predicted mean change to $\pi_{3(j)}$ (slope relating student grade level to Work Keys score) when traditional curricula are used rather than applied curricula.

$r_{3(j)}$ is a Level-2 random effect that represents the deviation of class (or grade) *j*'s slope coefficient (relating student grade level to Work Keys score) from its predicted value based on the Level-2 model. The random effects in these equations are assumed to be correlated. They are also assumed to be multivariate normally distributed with a mean of 0. The variance of this effect is designated as τ_{30} and has covariances τ_{31} and τ_{32} with the other Level-2 random effects.

APPENDIX C: DATA SET GRAPHS

OVERVIEW

Appendix C contains a complete set of graphs related to the data sets used in this investigation. As mentioned in Chapter 4, the original sample of 1,321 students resulted in full matrix data for 842 students after eliminating series with missing or obviously erroneous data points. Full matrix data for this investigation included information regarding school, type of course (applied or traditional), course, class within course, student gender, student grade in school, student cumulative grade point average (GPA), Iowa Tests of Educational Development score (Iowa percentile rank), the name of the Work Keys Assessment test, and the Work Keys Assessment test score. The remaining 479 students had missing data in one or more of the full matrix data series variables. It is not always necessary to restrict the investigation to those students included in the full matrix data series. Often it makes sense to use the entire student sample providing data on one specific variable even if they were unable to provide data on another. For example, one can look at the distribution of all students providing GPAs even if some of them did not take the Iowa Tests of Educational Development (ITED). When used, this data will be identified as vector data as opposed to the full matrix data. As one can ascertain from Table 4.2, the most common missing independent data were ITED scores; ITED are not universally required and many students simply do not take them.

The following graphs include data on grade, gender, GPA, ITED score, and Work Keys score variables compiled over three levels--student, class, and school. Each of these variables compares "applied" students versus "traditional" students in specific subgroups. The first series of Level 1 histograms include all students enrolled in applied courses versus all students enrolled in traditional courses; followed by histograms of students enrolled in specific applied courses versus their counterparts in specific traditional courses. The second series of Level 1 histograms include data on grade, gender, GPA, ITED score, and Work Keys score variables compiled only for those students who scored below the minimum competency level on each of the three Work Keys tests included in the study. These graphs were made to allow examination of student characteristics for potential patterns in the important subgroup of students who did not meet minimum levels of "employability skills". The Level 2 graphs follow the same pattern as the first series of Level 1 graphs with the exception that class means are charted rather than individual student results. The y-axis labels for all histograms are "frequency"; for Level-1 histograms this means "Number of Students", for Level-2 histograms this means "Number of Classes".

Boxplots are used to compare school-level data collected on applied versus traditional student groups; that is, Level 3 data will be school means for all students enrolled in applied courses versus school means for all students enrolled in traditional courses.

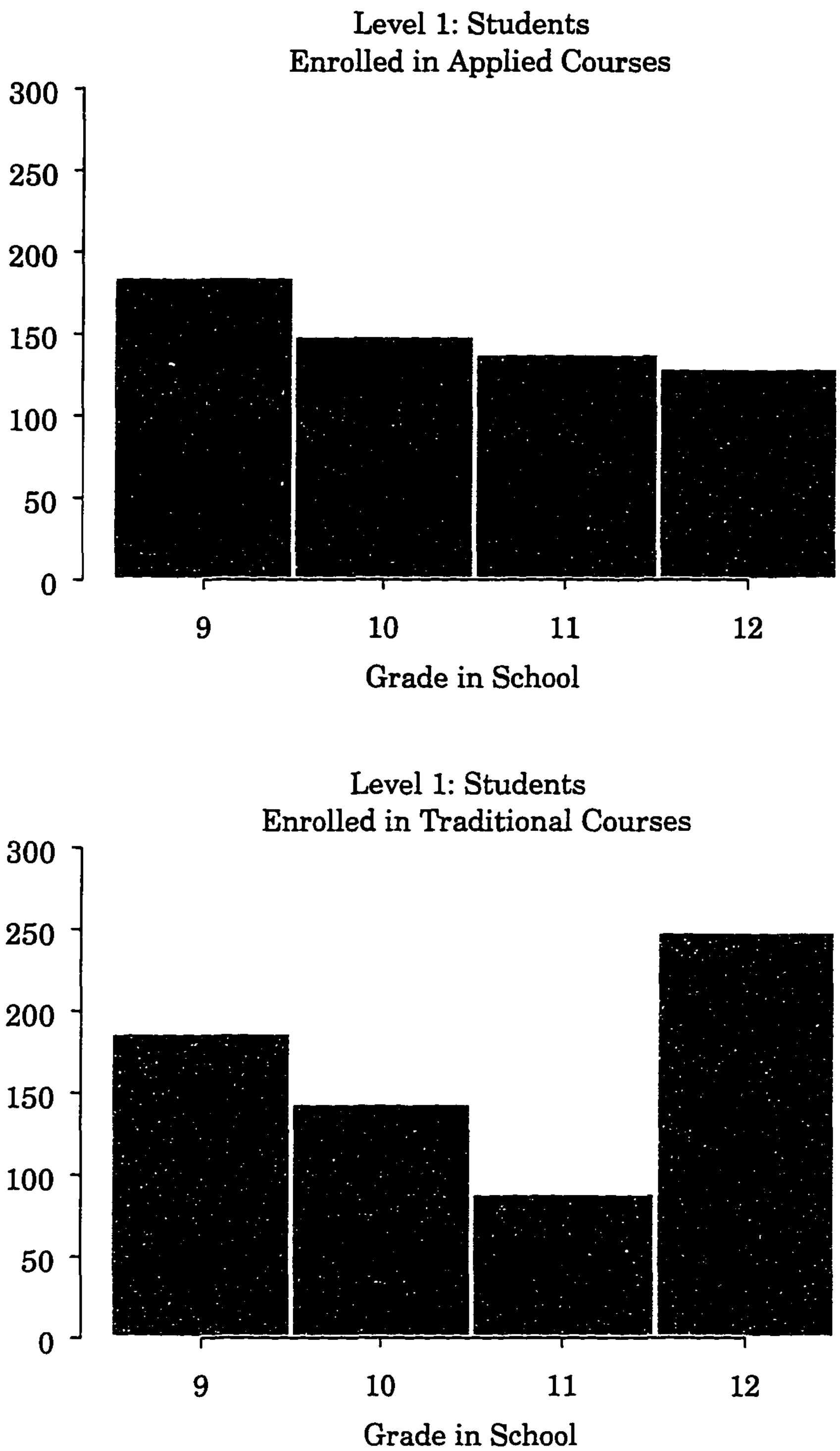
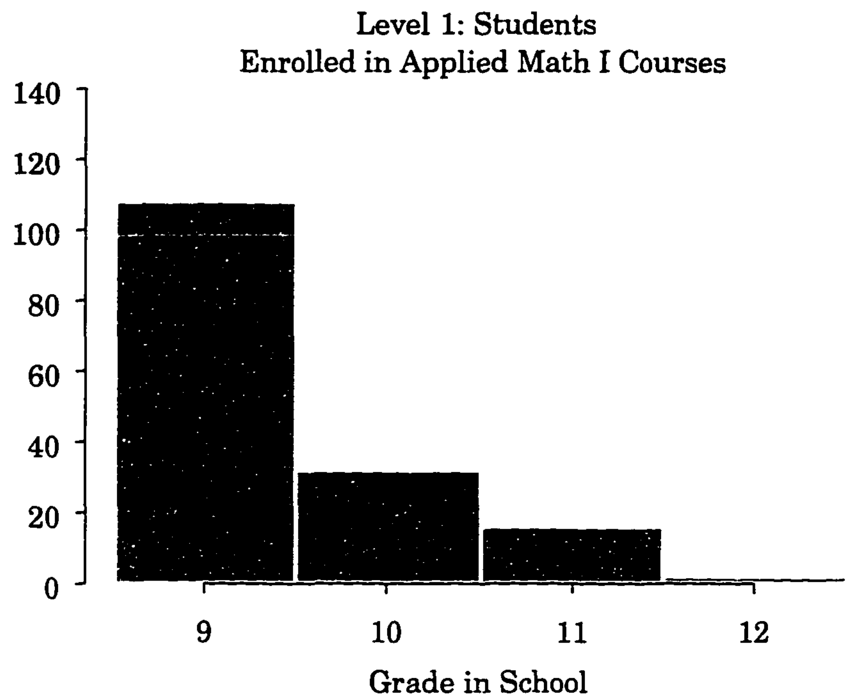


Figure C.1. Histograms comparing "applied" students' grade in school versus "traditional" students' grade in school (vector data)



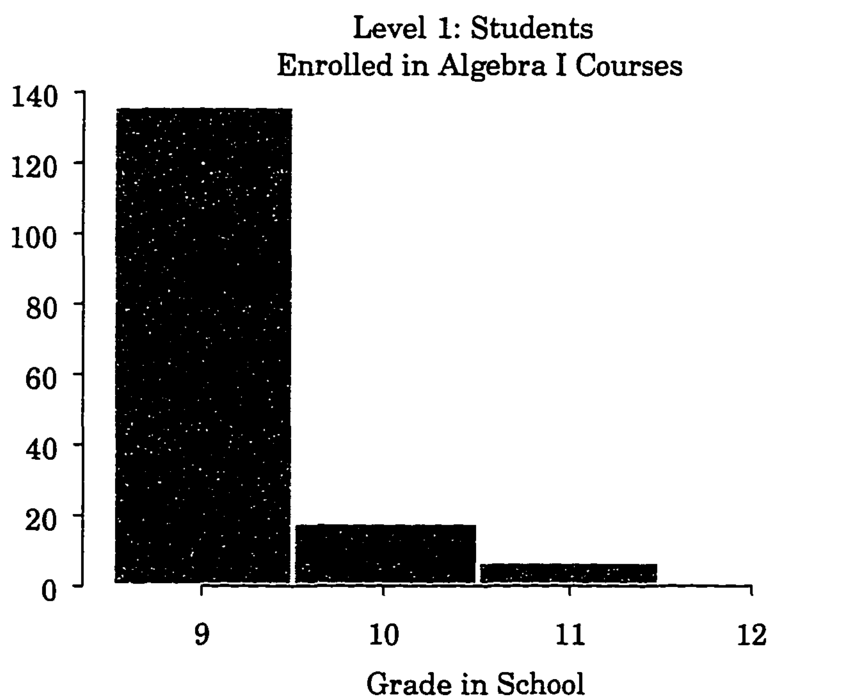


Figure C.2. Histograms comparing Applied Math I versus Algebra I students' grade in school (vector data)

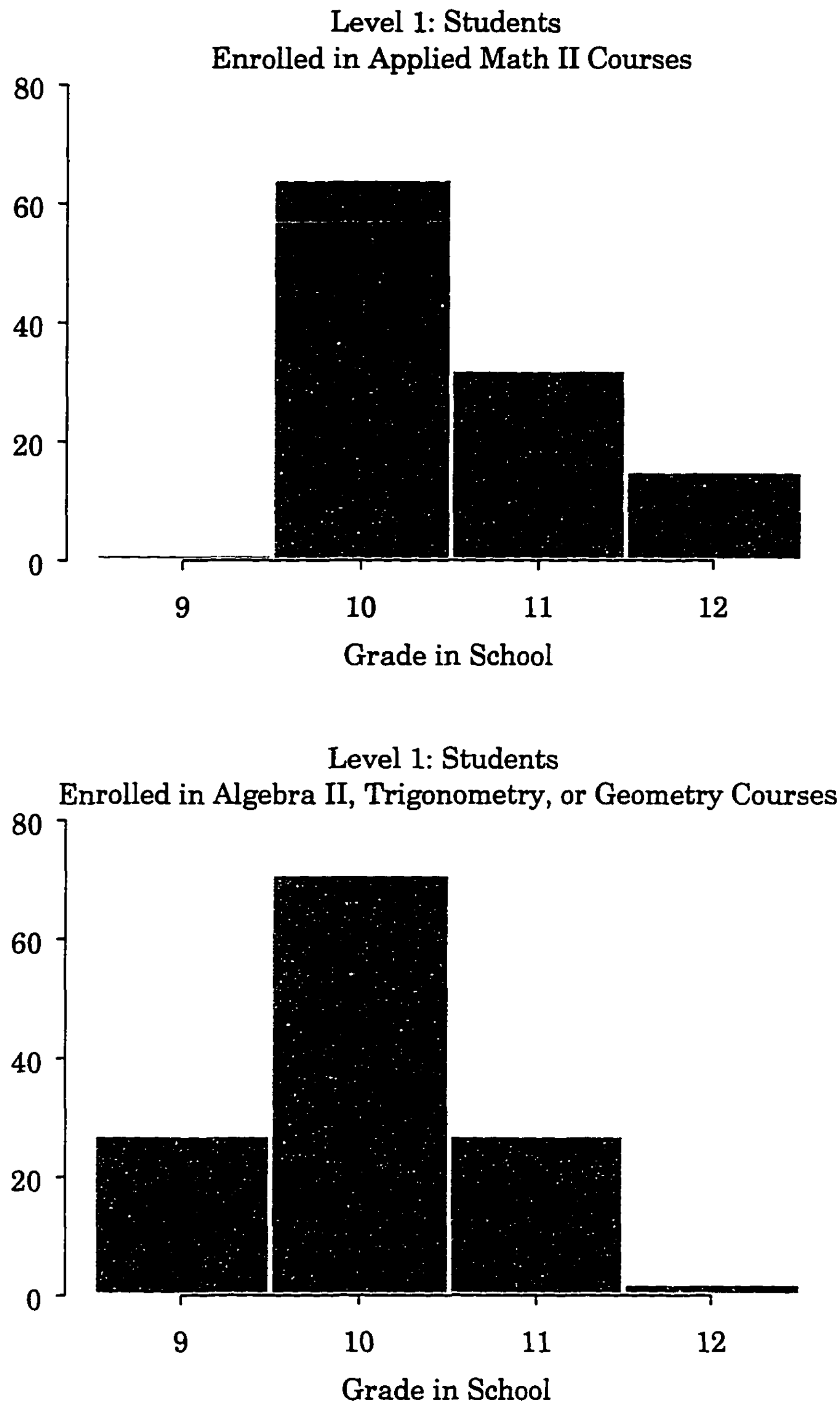


Figure C.3. Histograms comparing Applied Math II versus Traditional Math II students' grade in school (vector data)

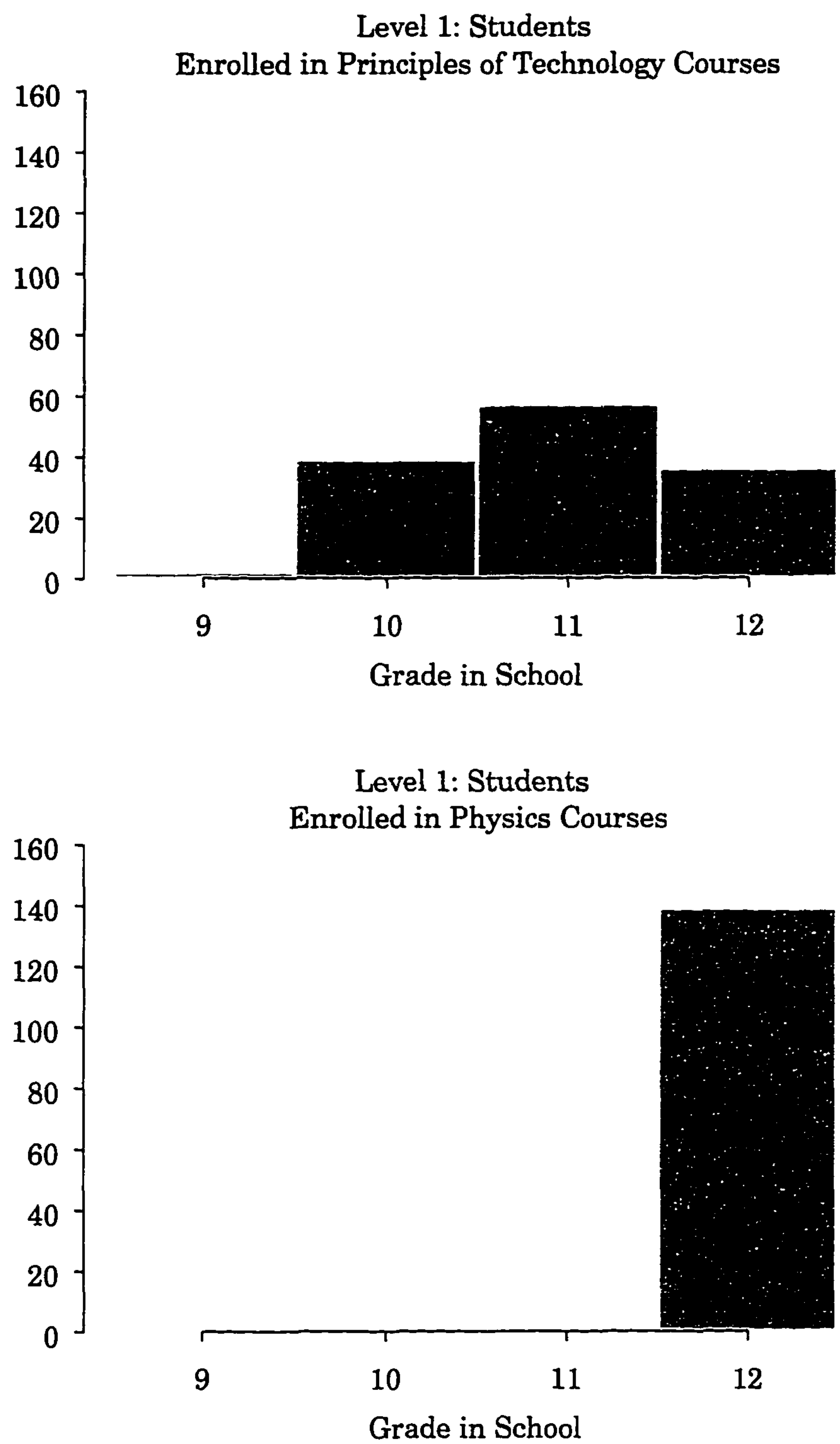
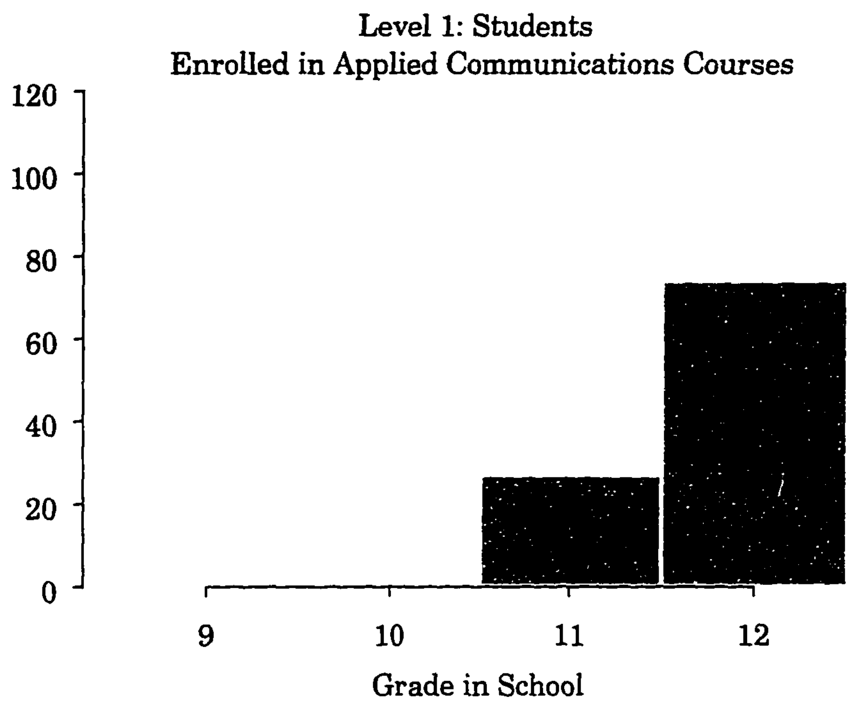


Figure C.4. Histograms comparing Principles of Technology students' grade in school versus Physics students' grade in school (vector data)



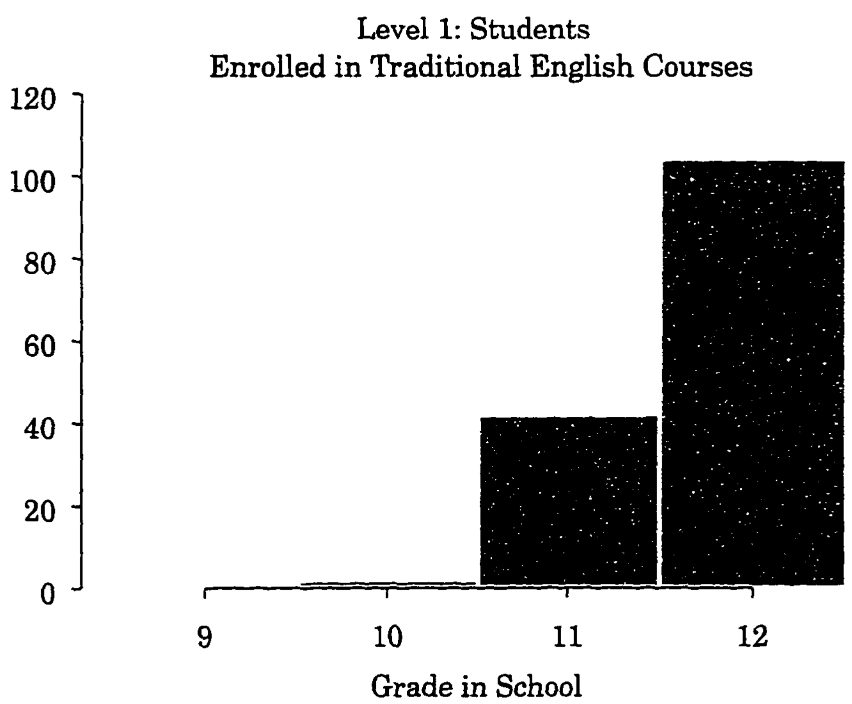
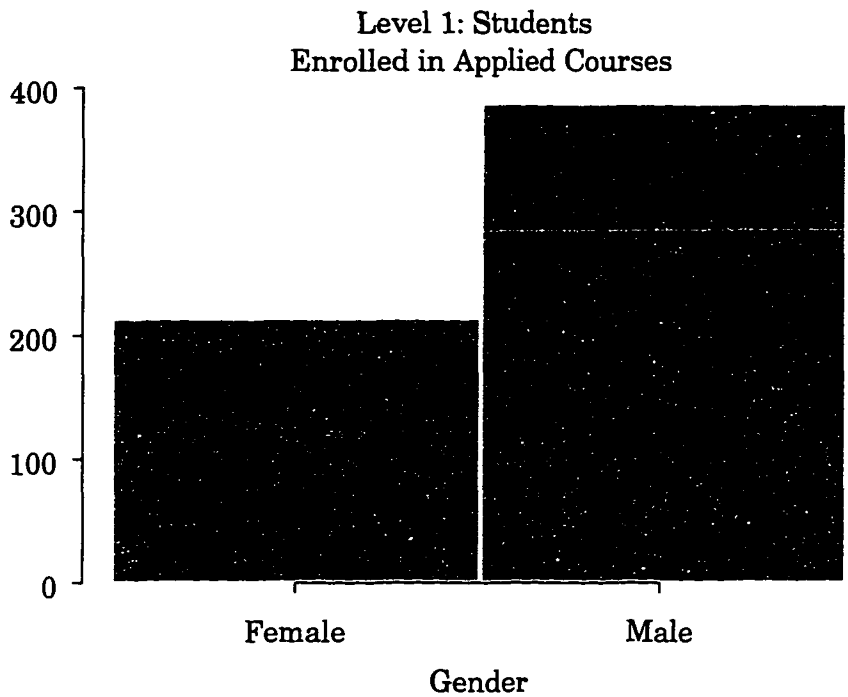


Figure C.5. Histograms comparing Applied Communications versus Traditional English students' grade in school (vector data)



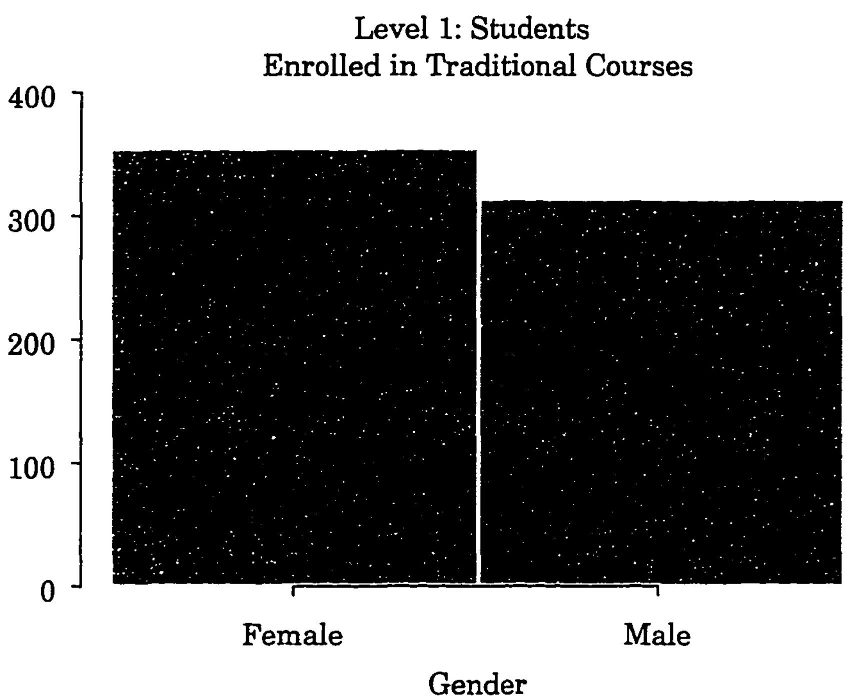
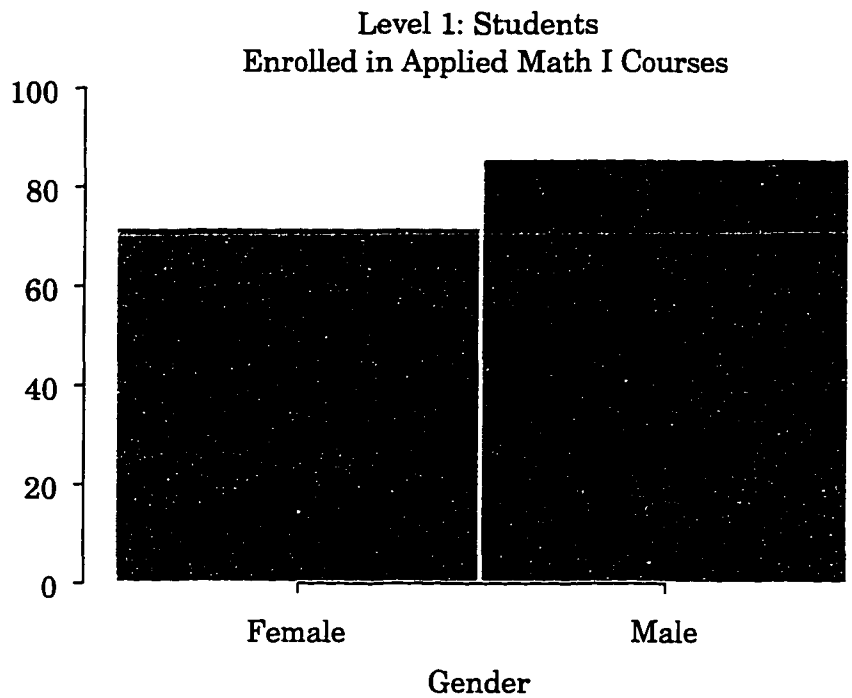


Figure C.6. Histograms comparing “applied” students’ gender ratio versus “traditional” students’ gender ratio (vector data)



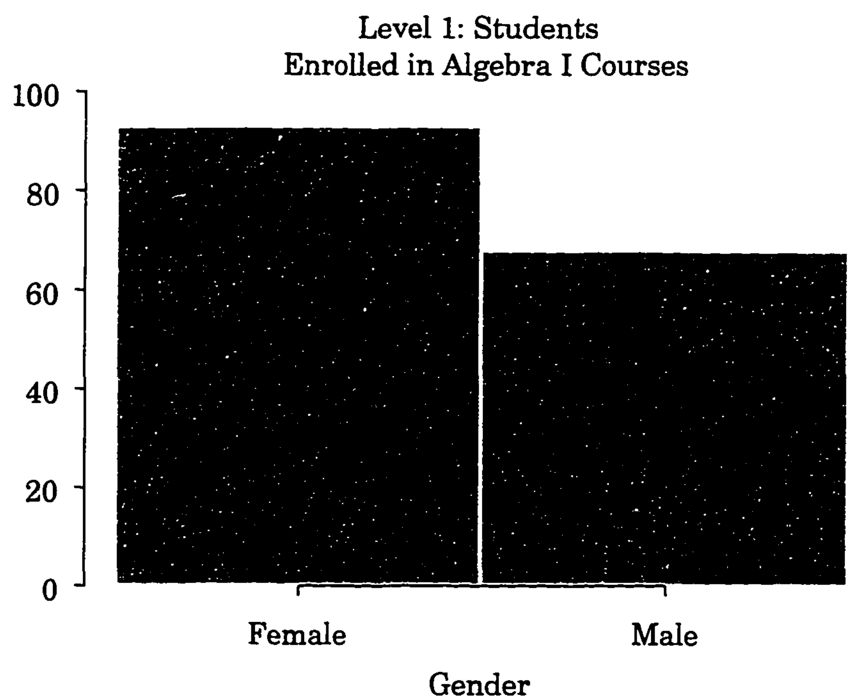
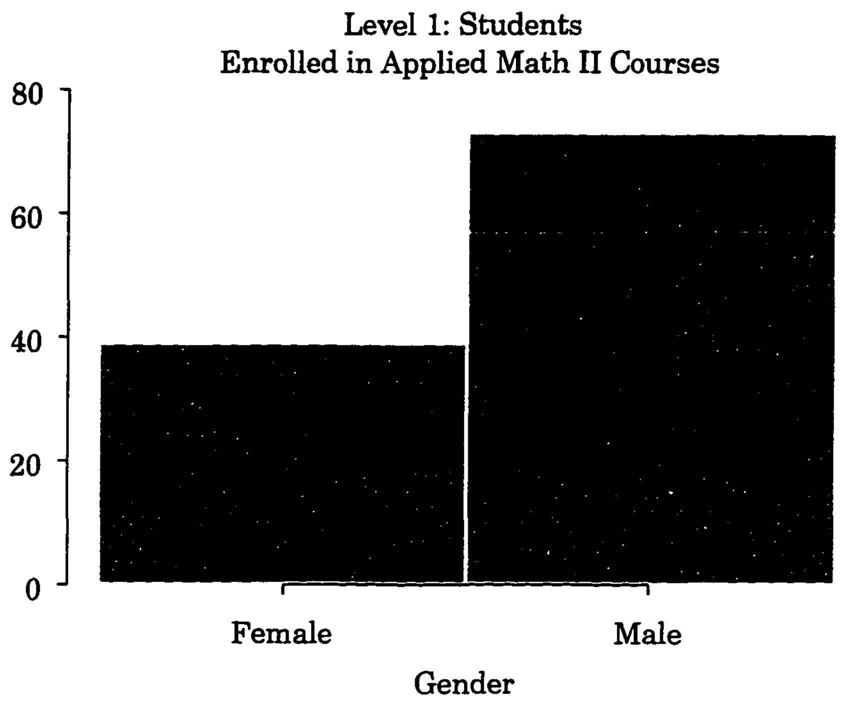


Figure C.7. Histograms comparing Applied Math I versus Algebra I students' gender ratio (vector data)



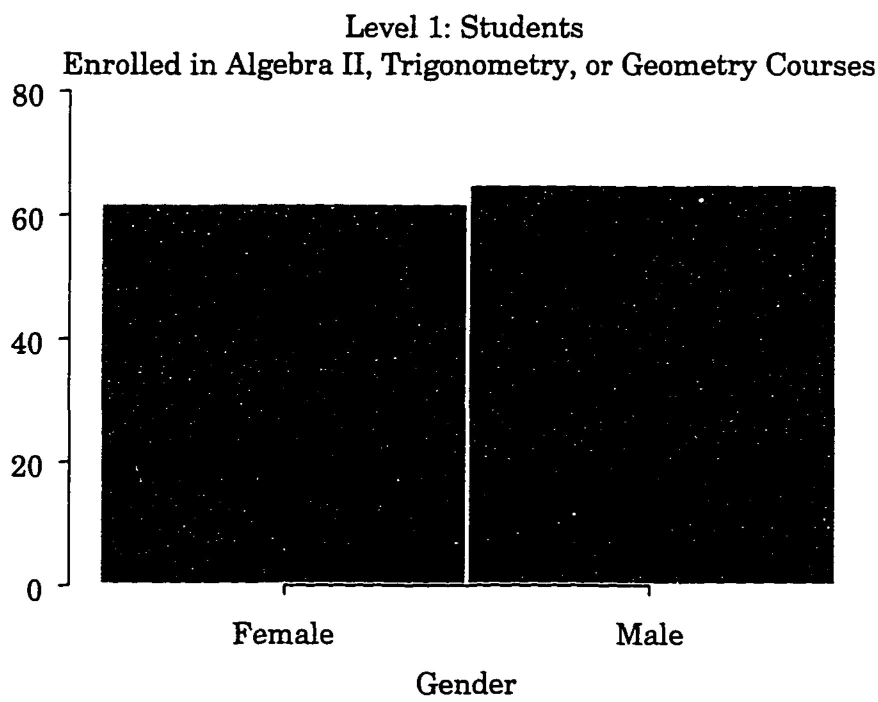
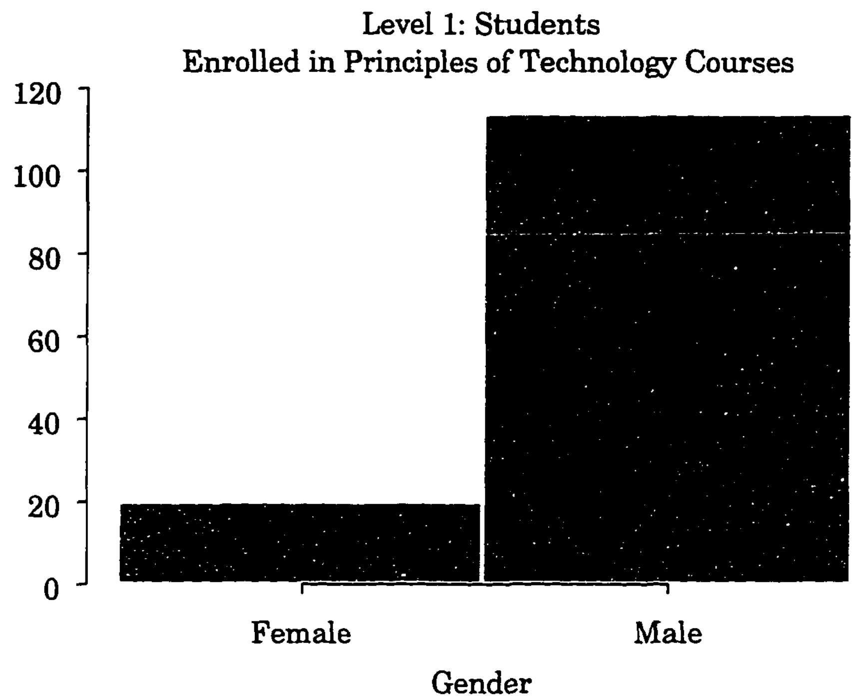


Figure C.8. Histograms comparing Applied Math II versus Traditional Math II students' gender ratio (vector data)



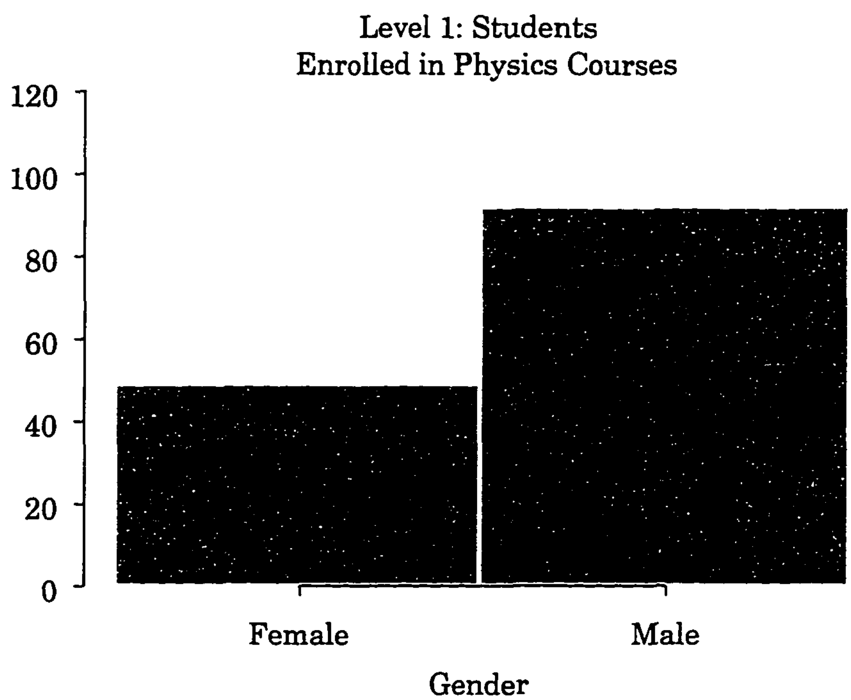
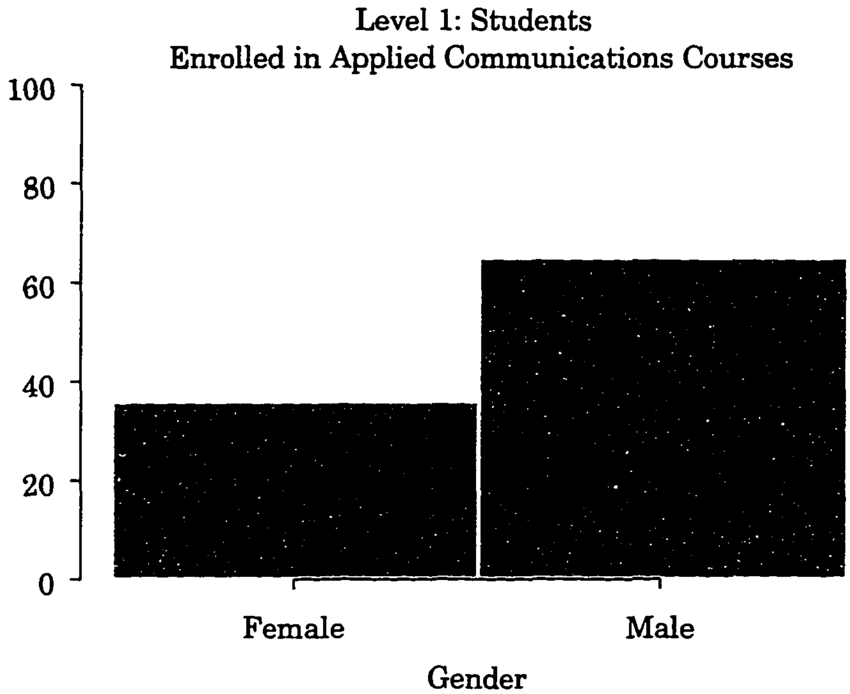


Figure C.9. Histograms comparing Principles of Technology versus Physics students' gender ratio in school (vector data)



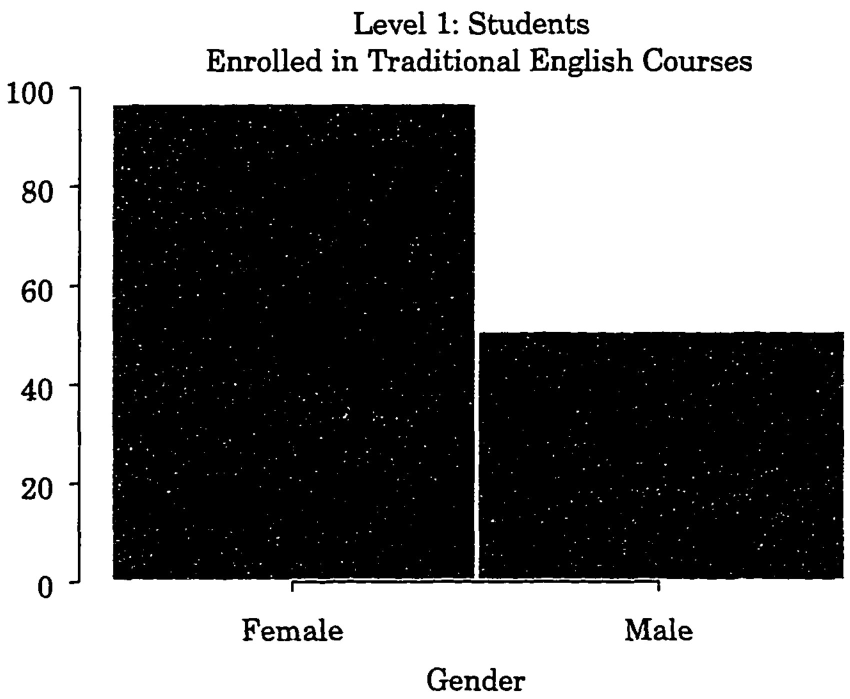


Figure C.10. Histograms comparing Applied Communications versus Traditional English students' gender ratio (vector data)

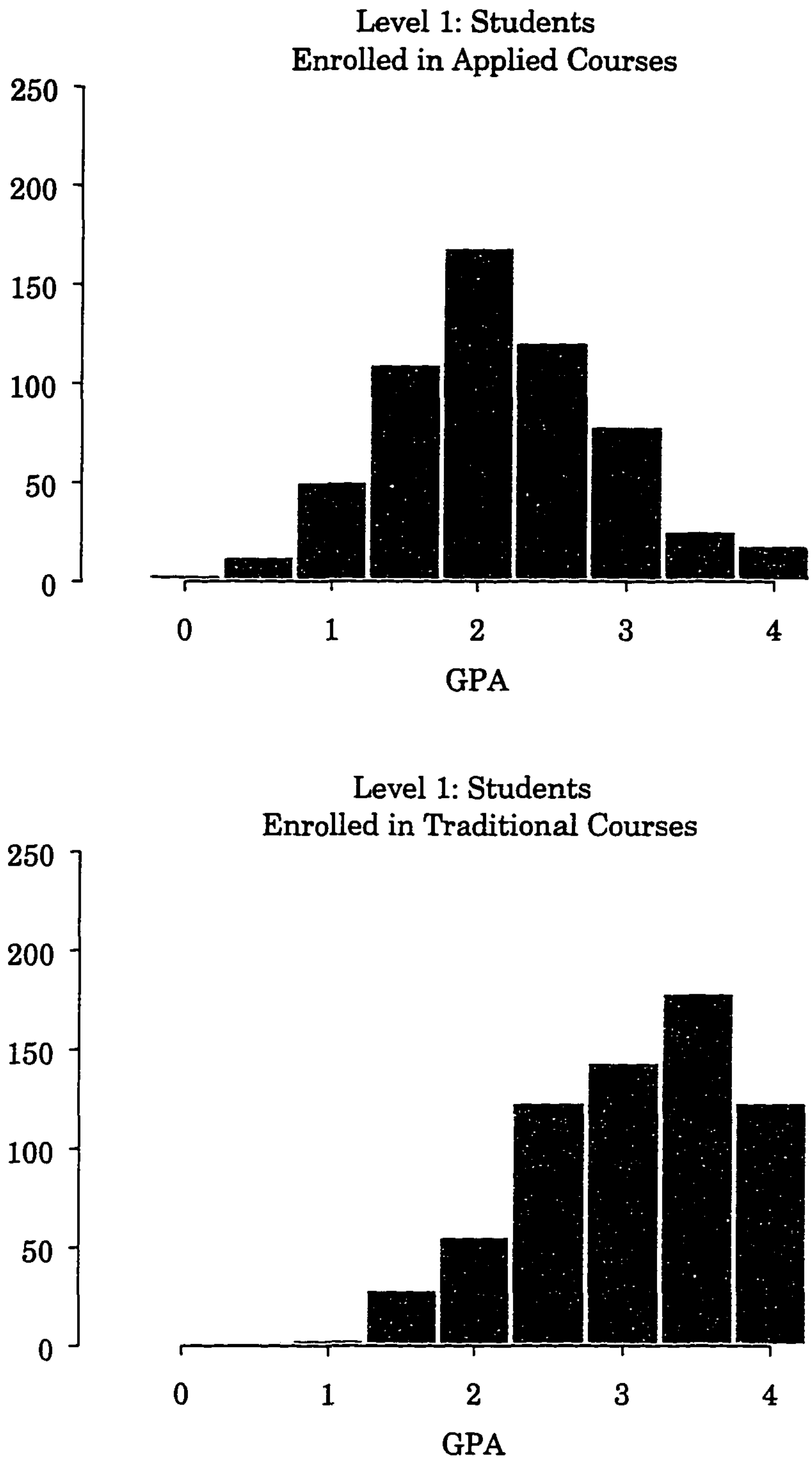


Figure C.11. Histograms comparing "applied" versus "traditional" students' grade point average (vector data)

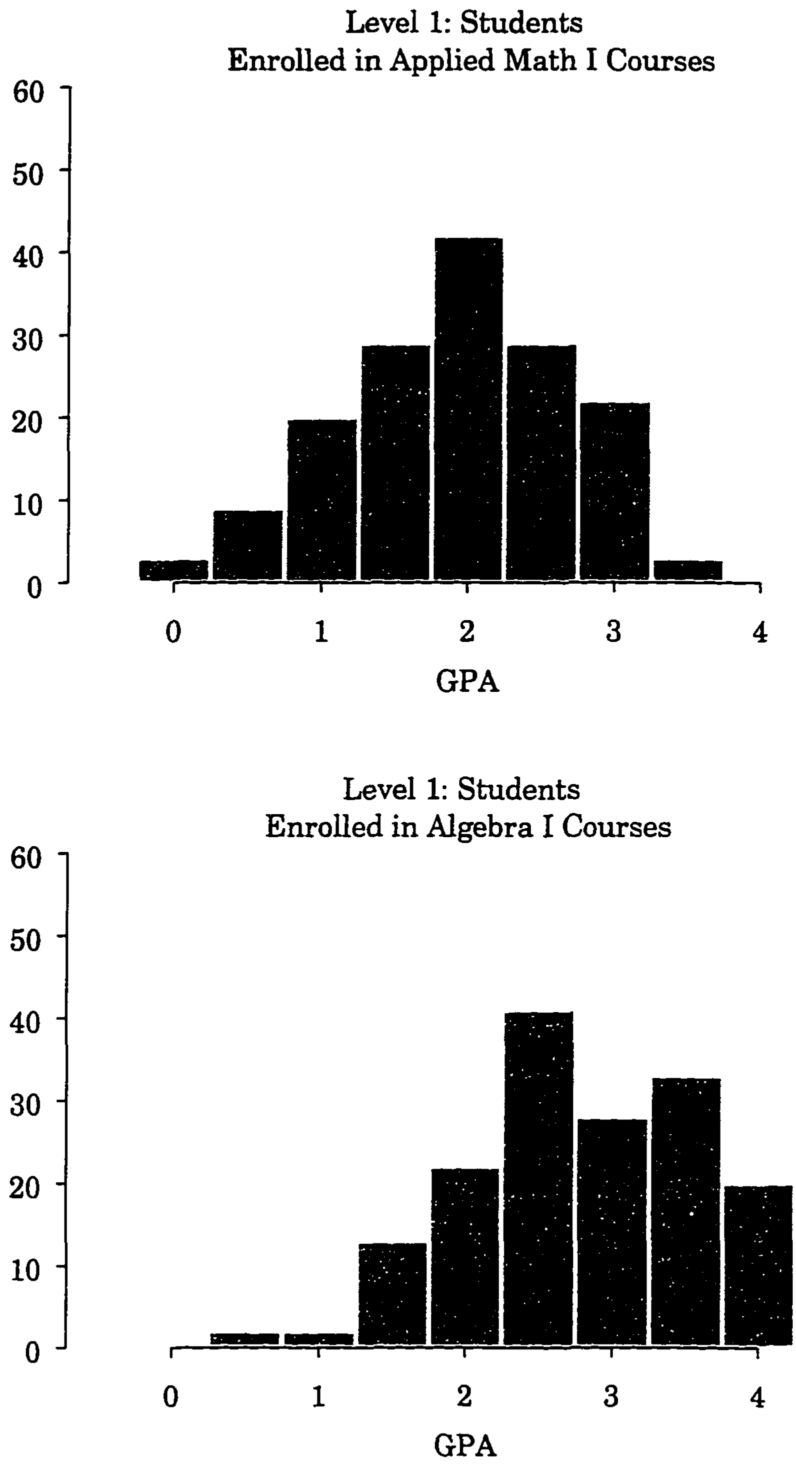
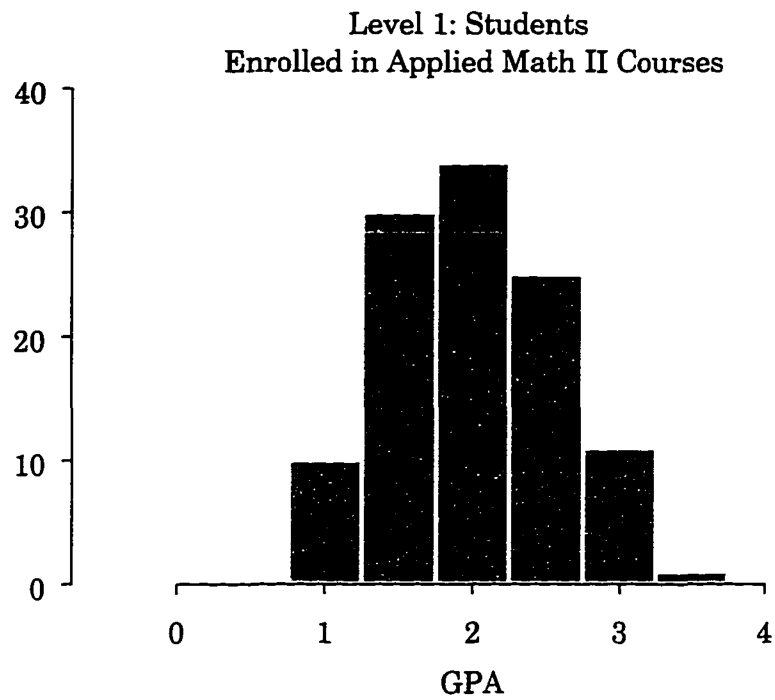


Figure C.12. Histograms comparing Applied Math I versus Algebra I students' grade point average (vector data)



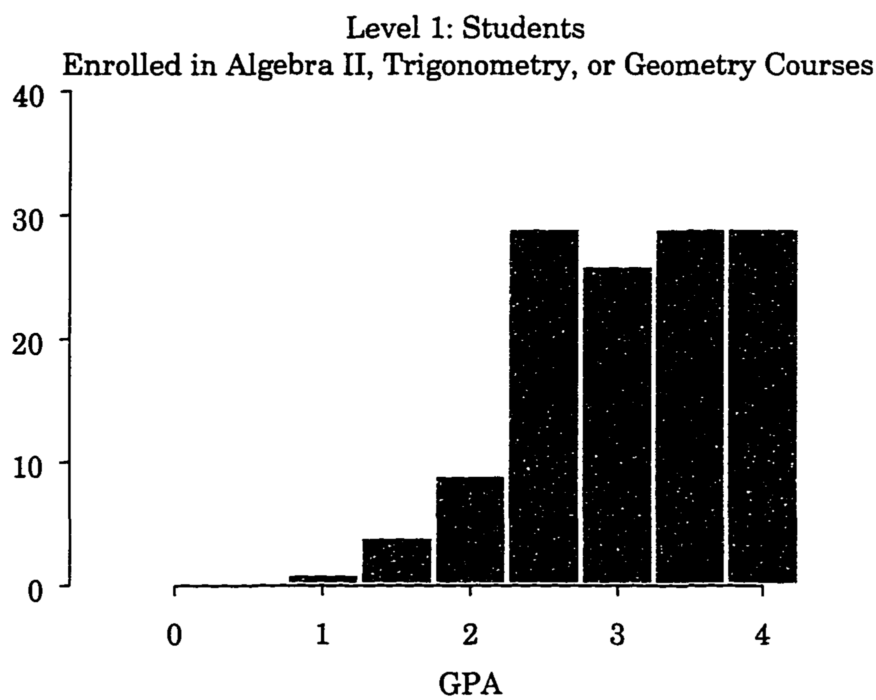
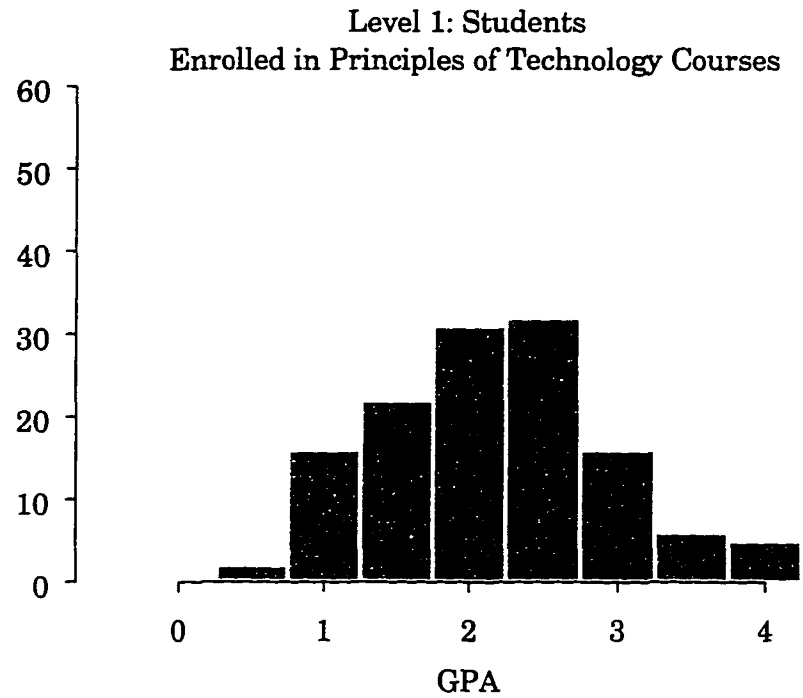


Figure C.13. Histograms comparing Applied Math II versus Traditional Math II students' grade point average (vector data)



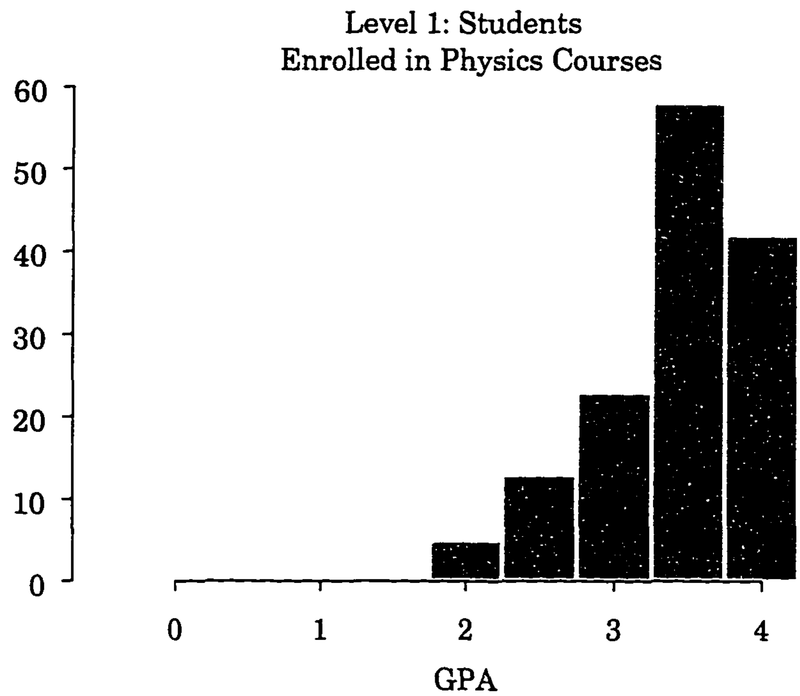


Figure C.14. Histograms comparing Principles of Technology versus Physics students' GPA (vector data)

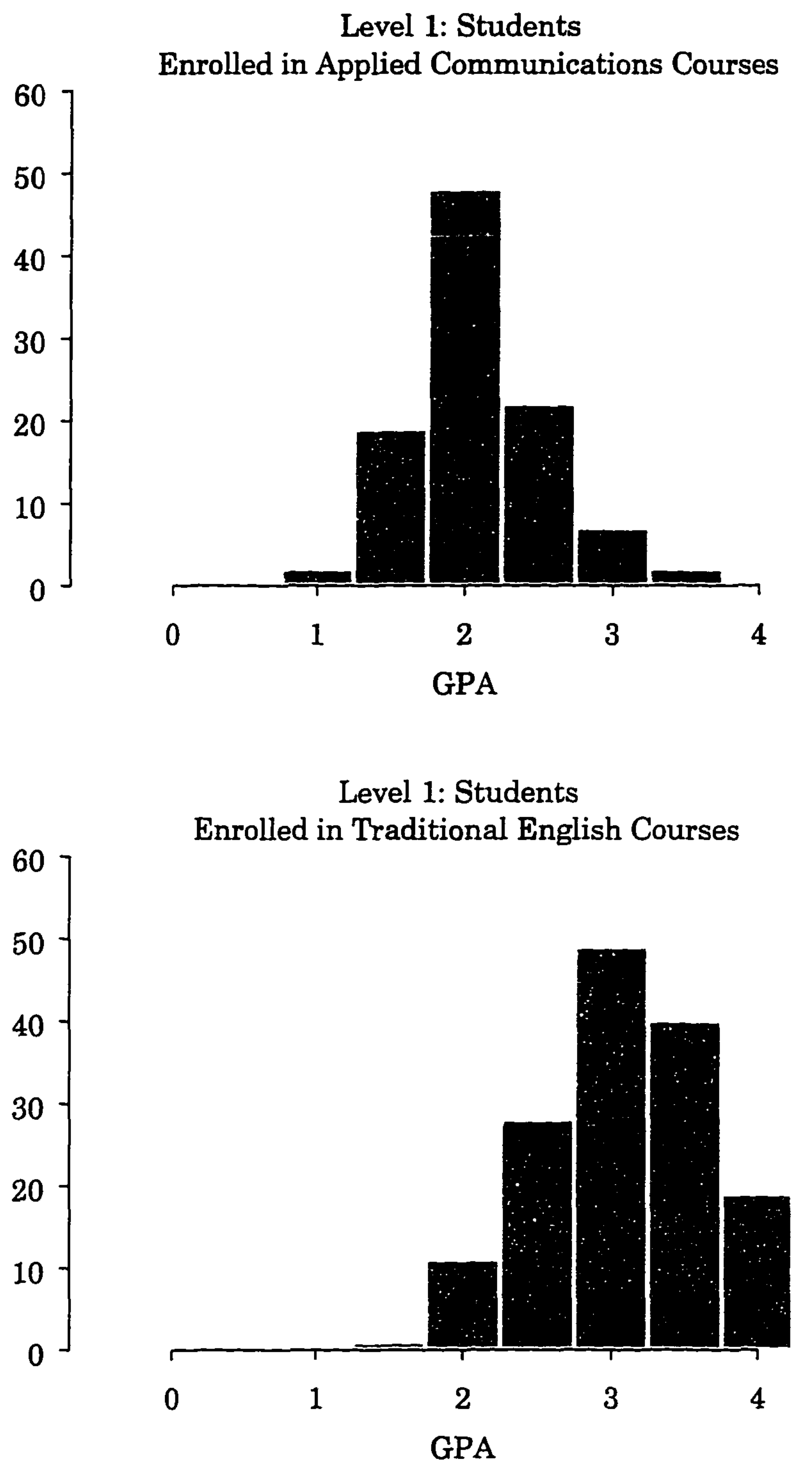


Figure C.15. Histograms comparing Applied Communications versus Traditional English students' grade point average (vector data)

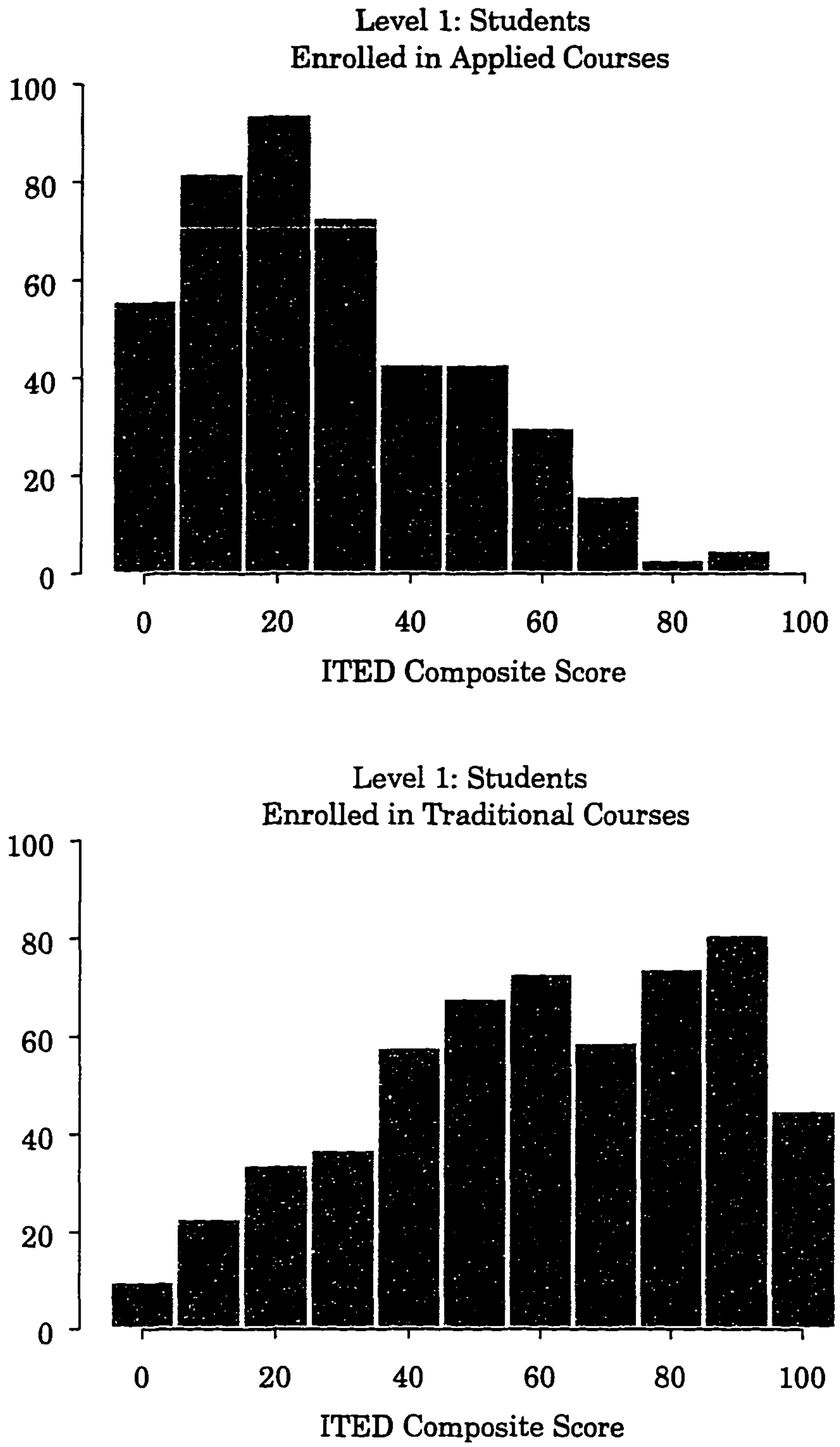
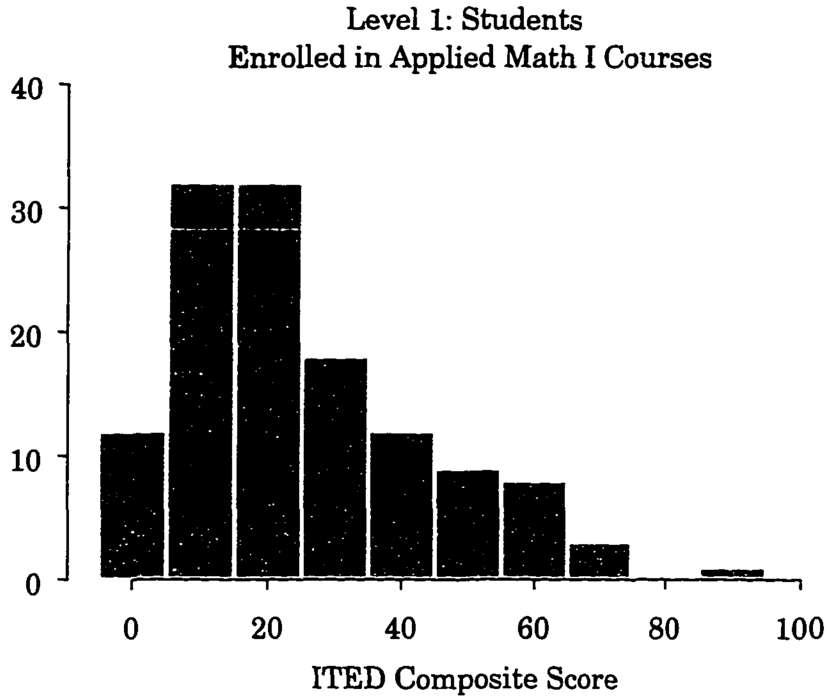


Figure C.16. Histograms comparing “applied” versus “traditional” students' ITED score (vector data)



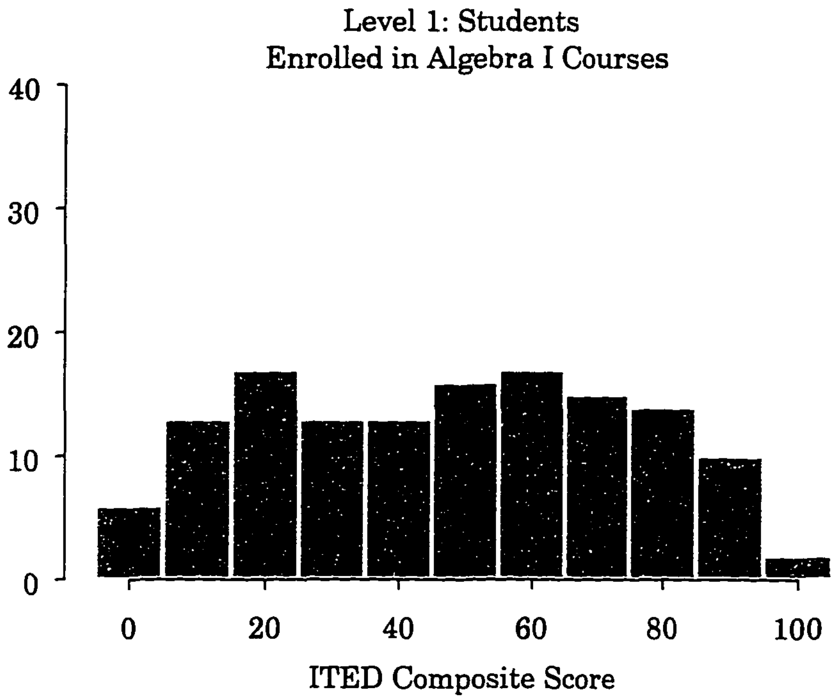
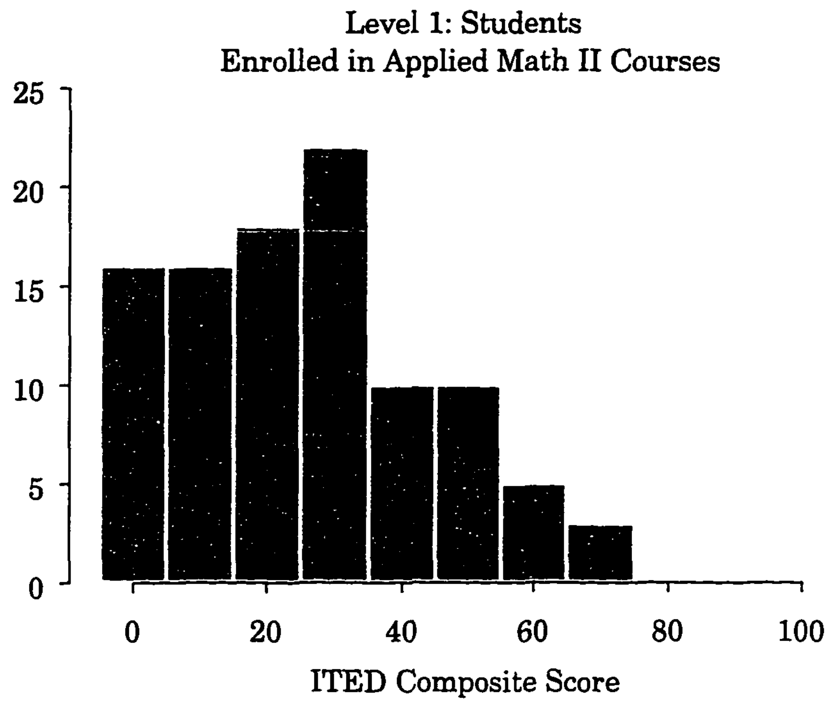


Figure C.17. Histograms comparing Applied Math I versus Algebra I students' ITED score (vector data)



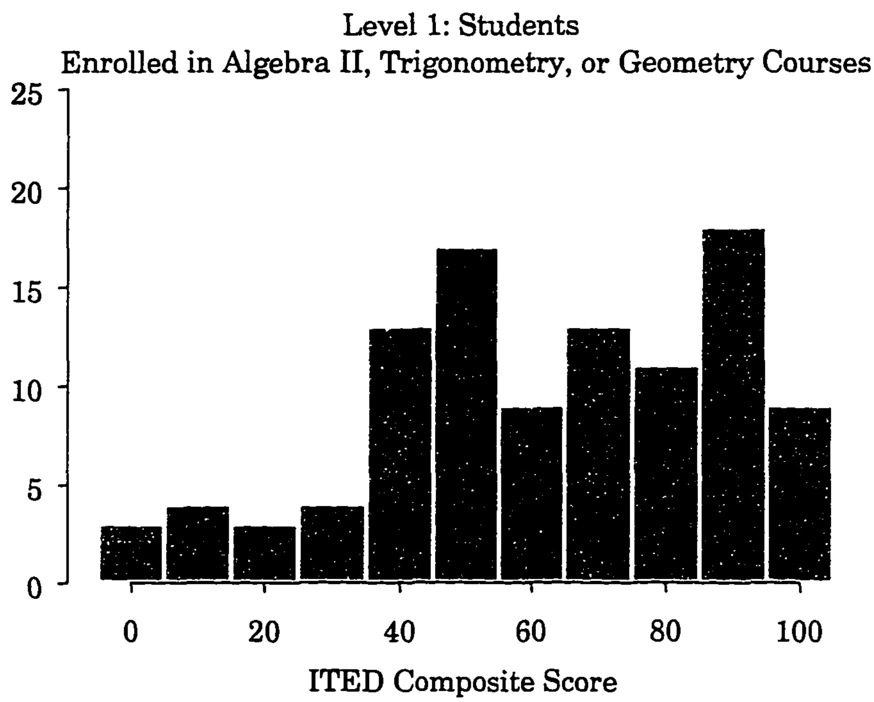
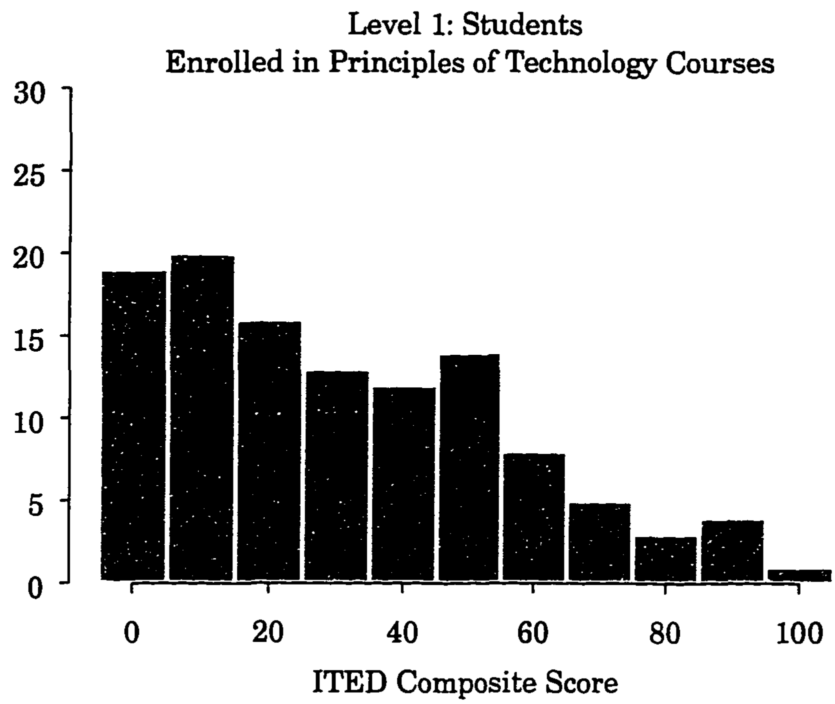


Figure C.18. Histograms comparing Applied Math II versus Traditional Math II students' ITED score (vector data)



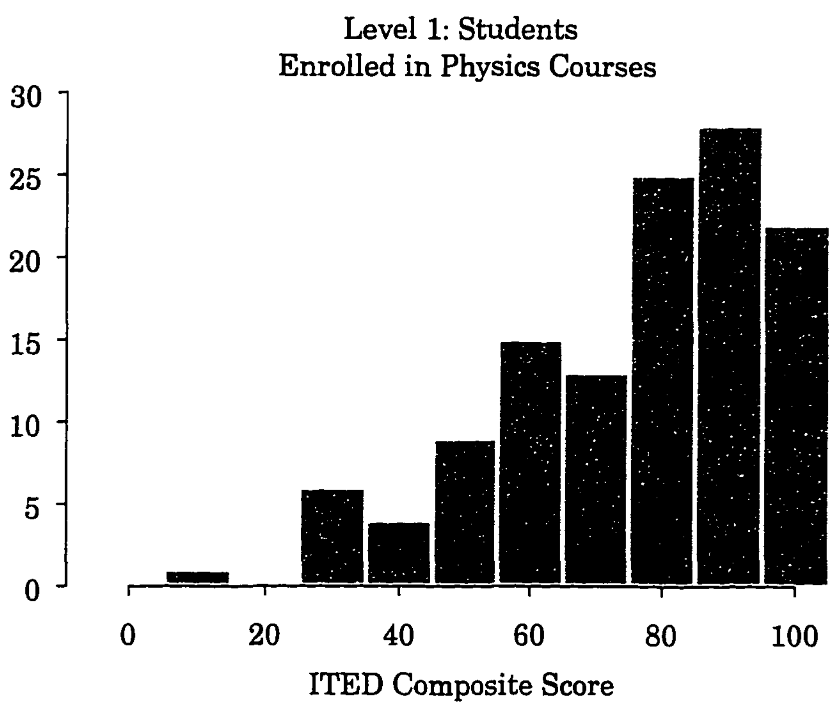


Figure C.19. Histograms comparing Principles of Technology versus Physics students' ITED score (vector data)

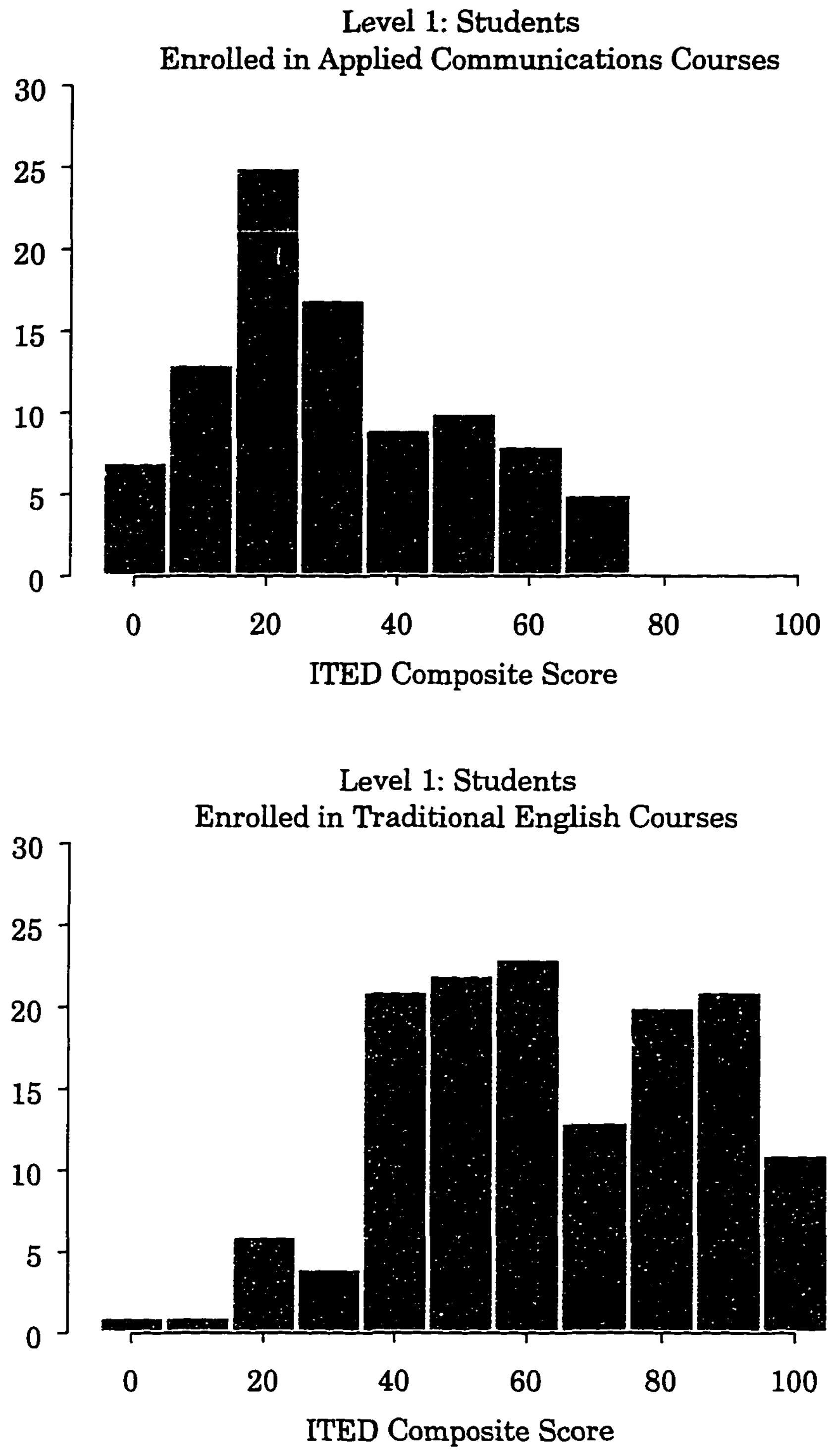
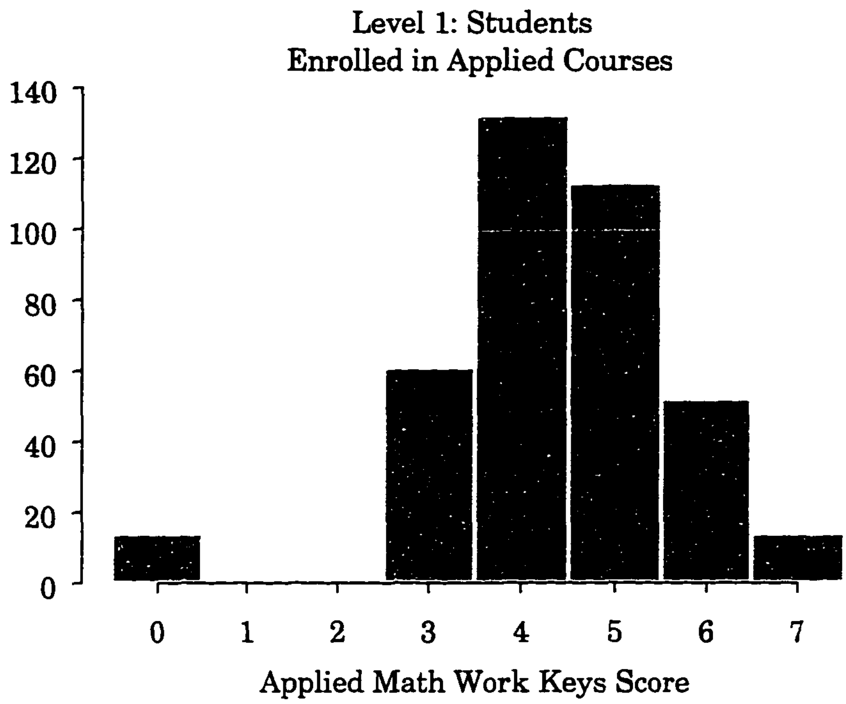


Figure C.20. Histograms comparing Applied Communications versus Traditional English students' ITED score (vector data)



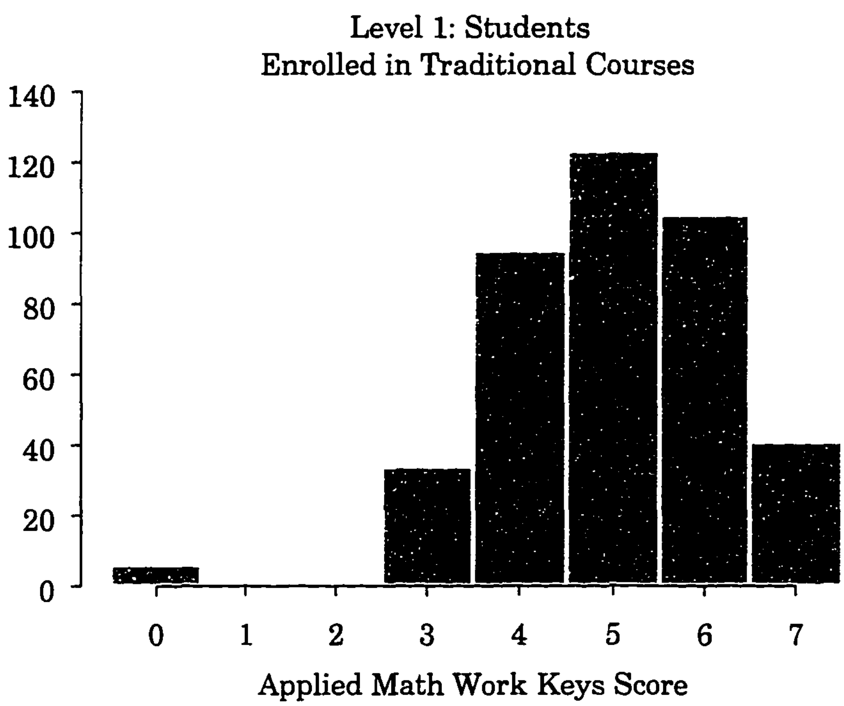
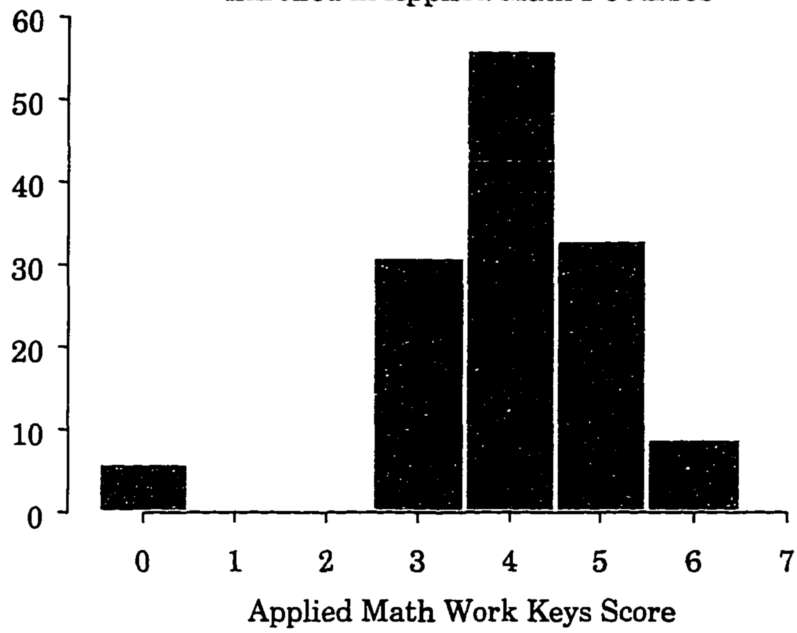


Figure C.21. Histograms comparing "applied" versus "traditional" students' Applied Math Work Keys score (vector data)

Level 1: Students
Enrolled in Applied Math I Courses



Level 1: Students
Enrolled in Algebra I Courses

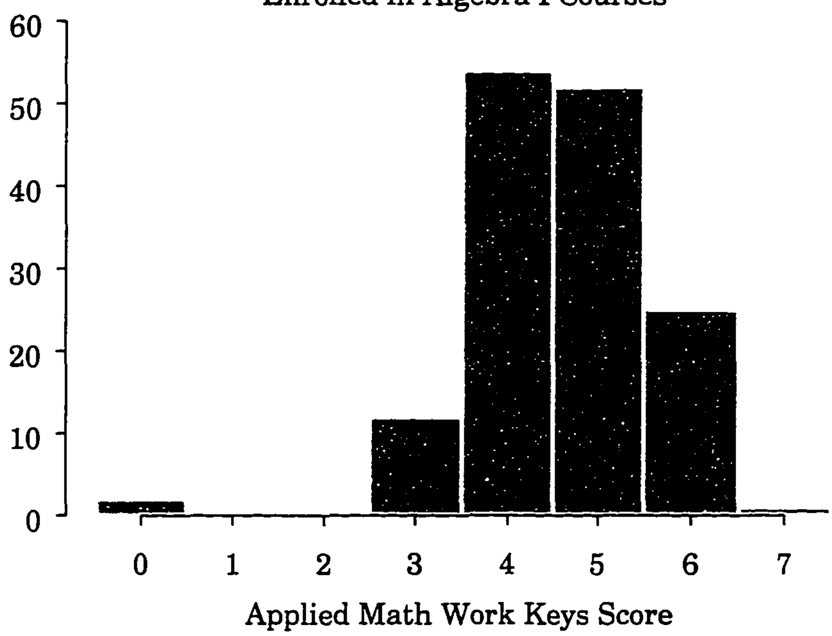


Figure C.22. Histograms comparing Applied Math I versus Algebra I students' Applied Math Work Keys test score (vector data)

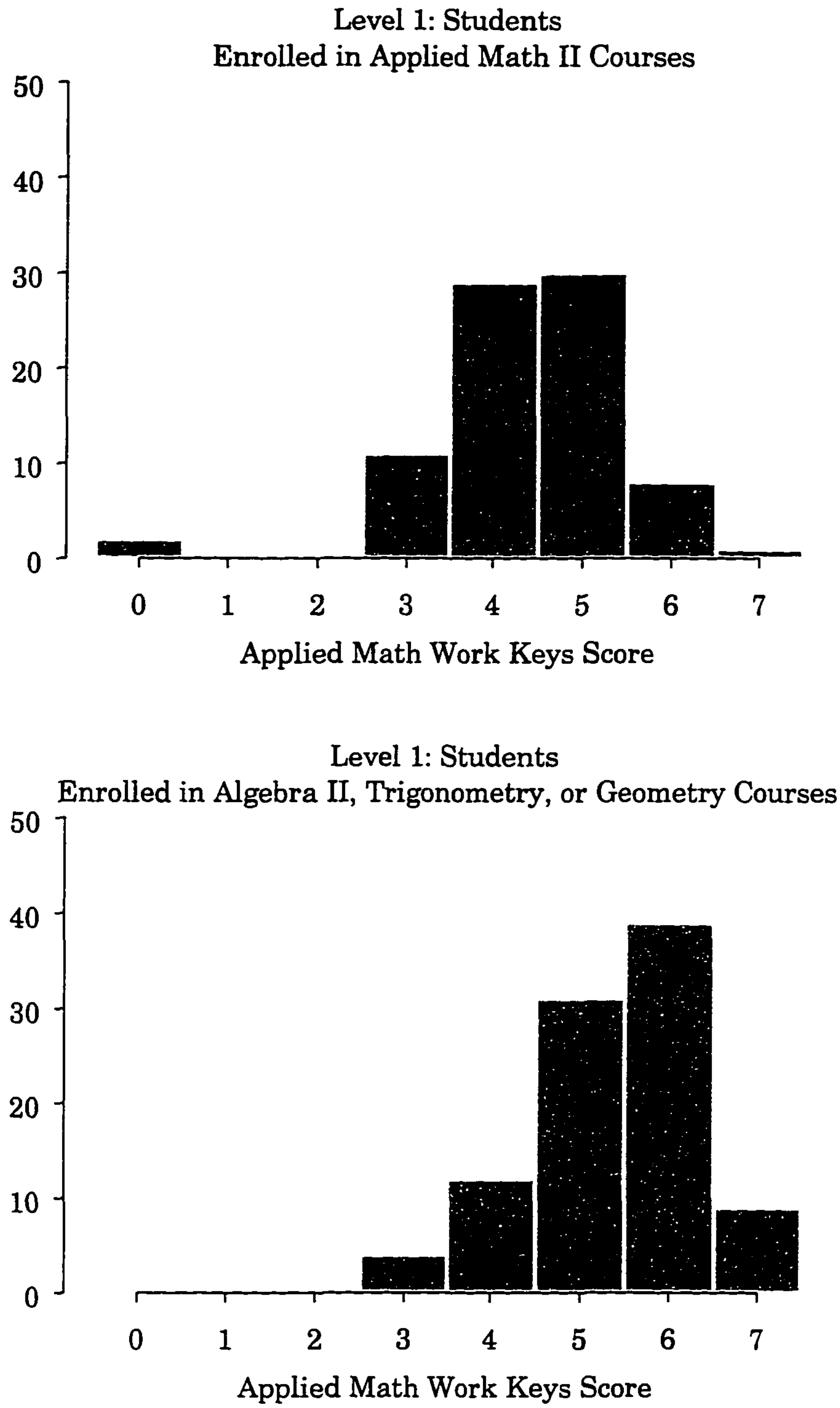


Figure C.23. Histograms comparing Applied Math II versus Traditional Math II students' Applied Math Work Keys Score (vector data)

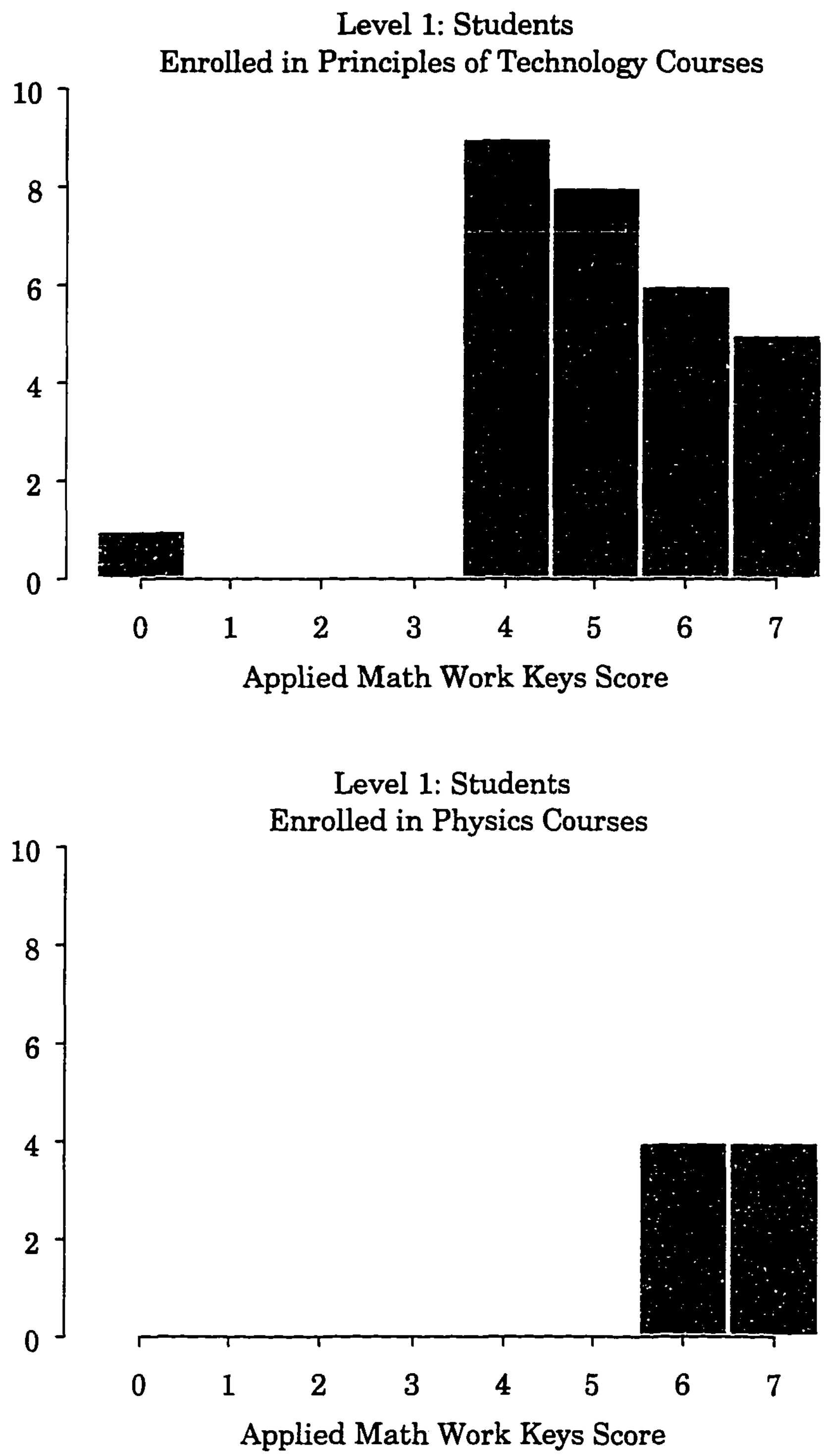
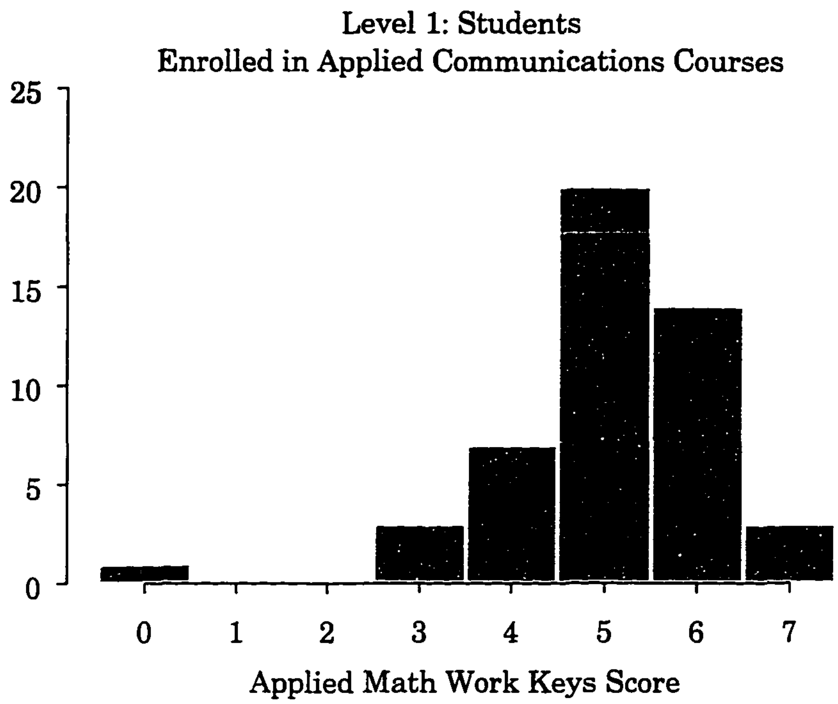


Figure C.24. Histograms comparing Principles of Technology versus Physics students' Applied Math Work Keys score (vector data)



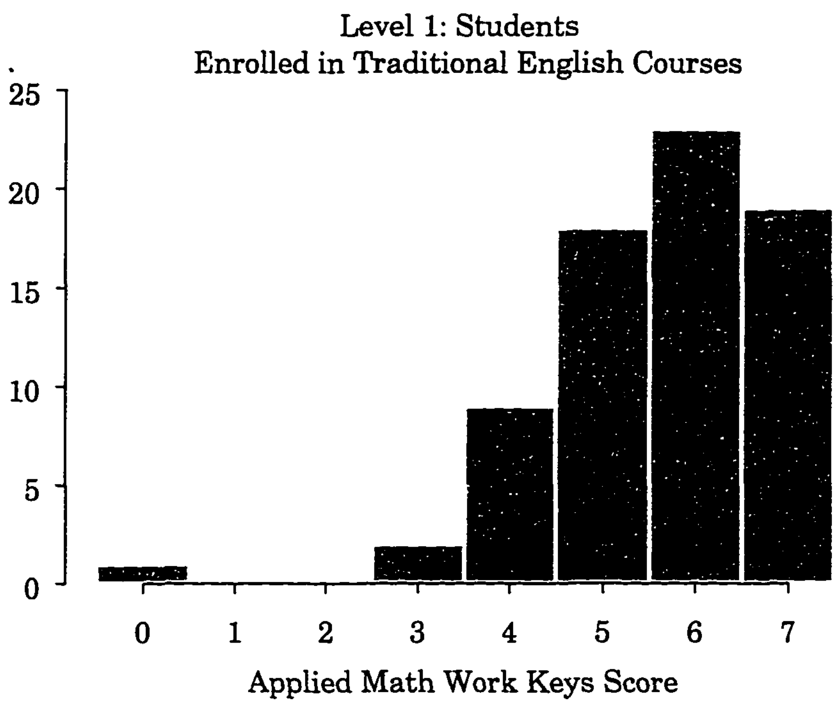
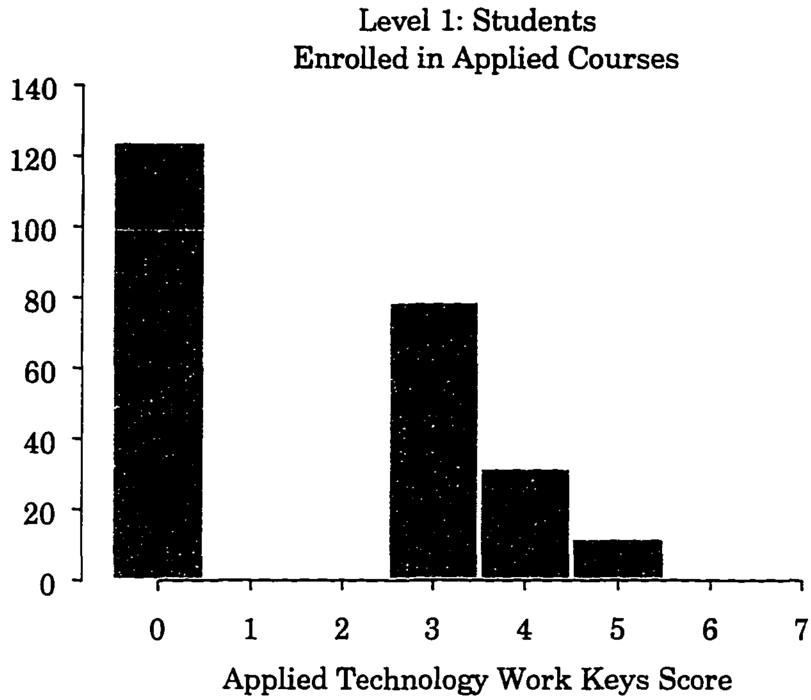


Figure C.25. Histograms comparing Applied Communications vs. Traditional English students' Applied Math Work Keys score (vector data)



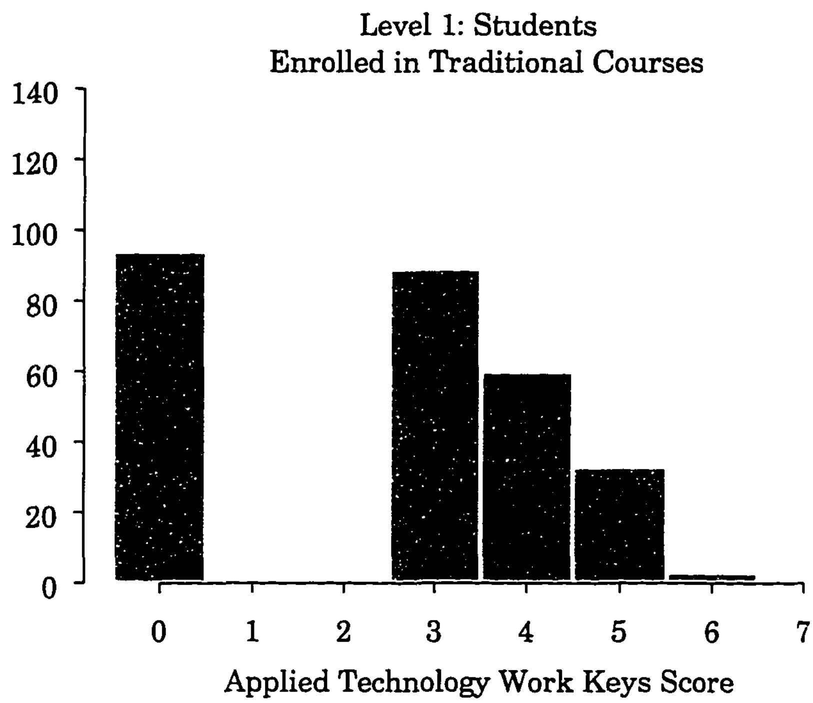


Figure C.26. Histograms comparing "applied" versus "traditional" students' Applied Technology Work Keys score (vector data)

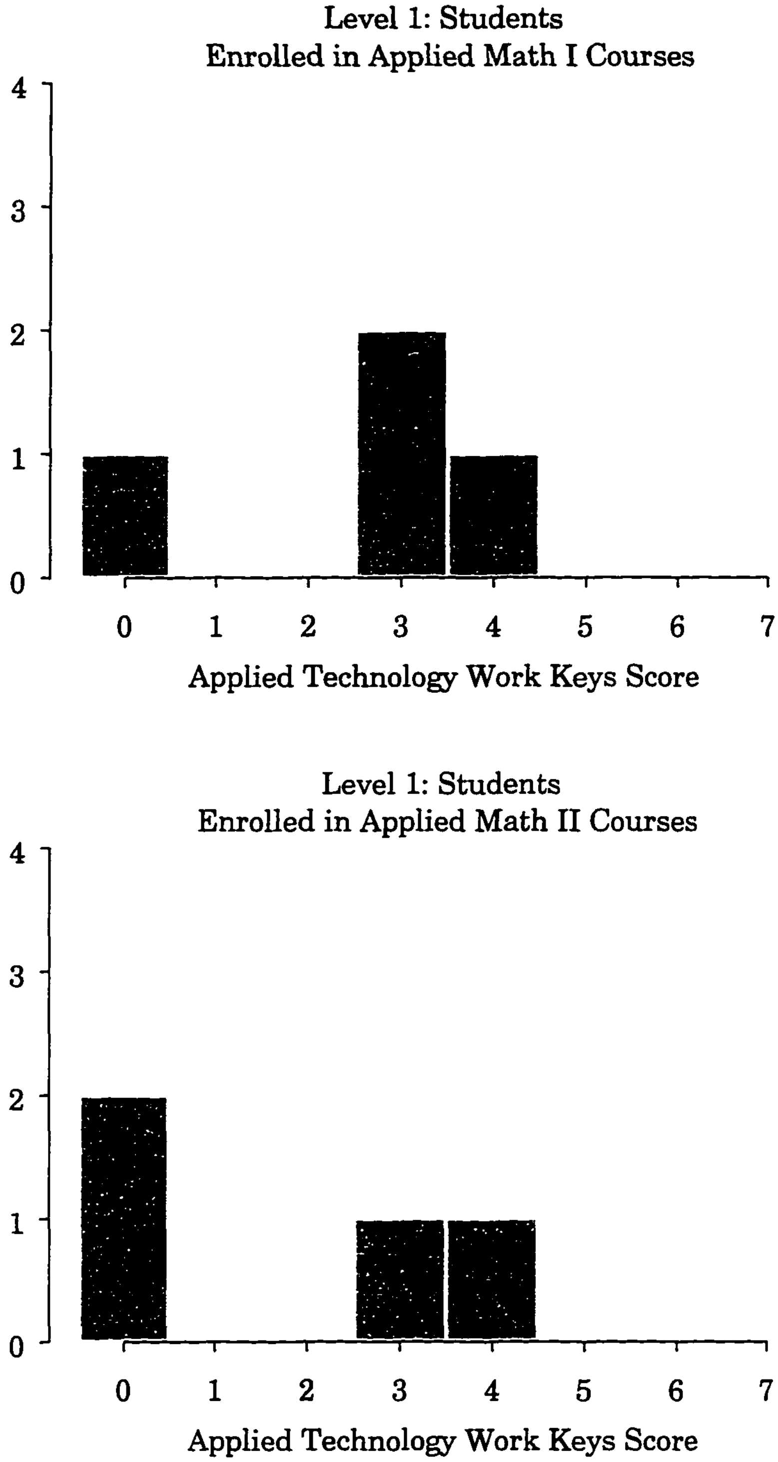
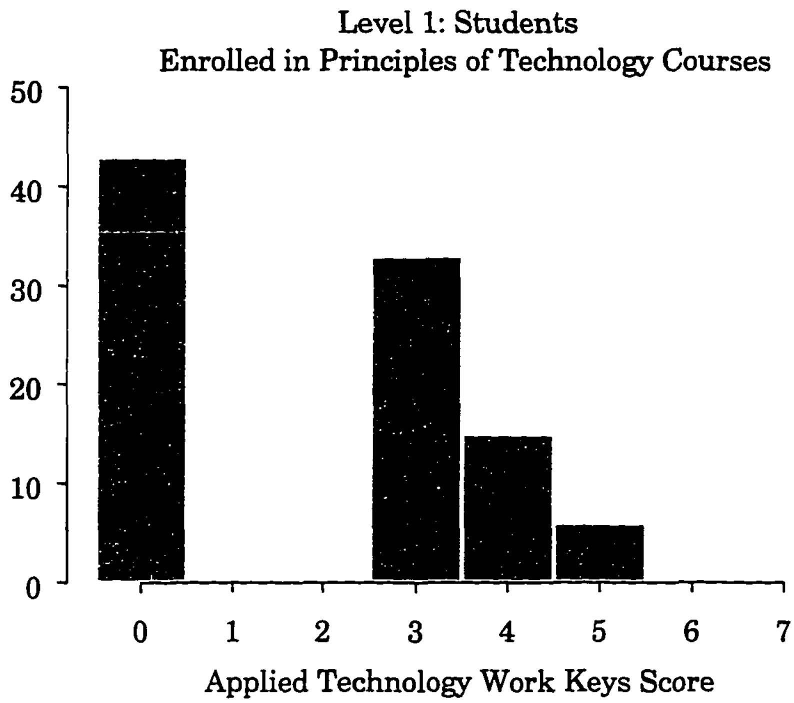


Figure C.27. Histograms of Applied Math I and II students' Applied Technology Work Keys score (vector data), no Traditional Math data available



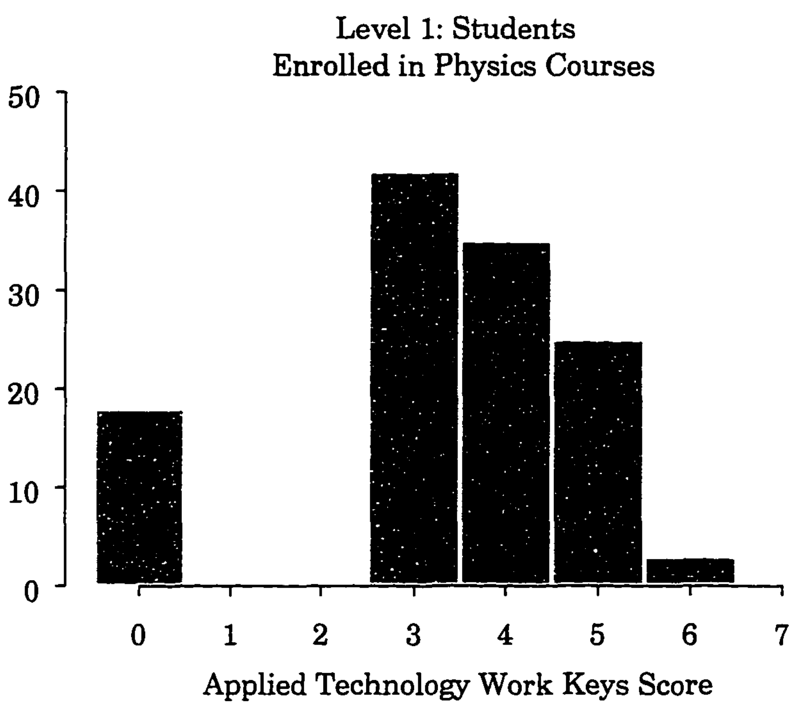
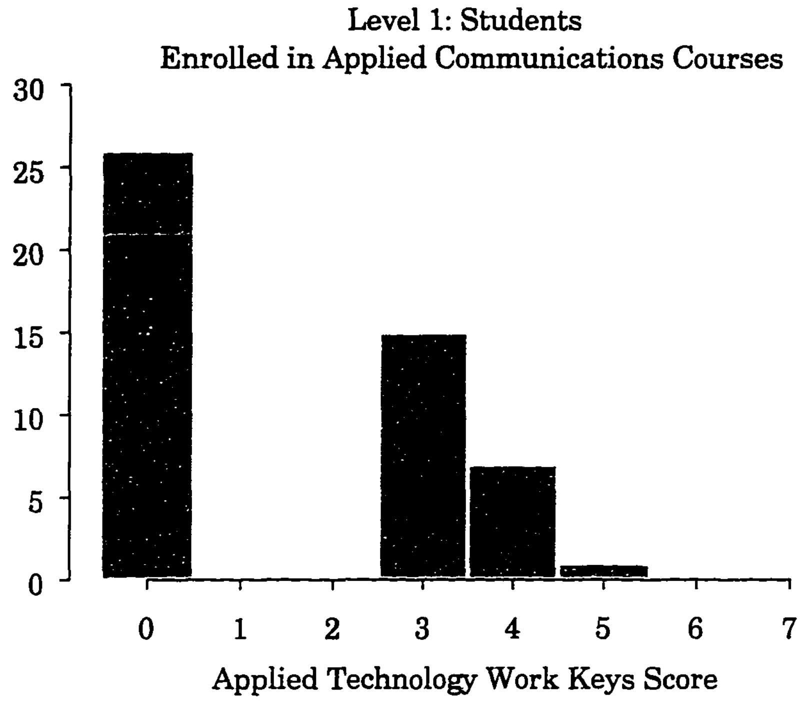


Figure C.28. Histograms comparing Principles of Technology versus Physics students' Applied Technology Work Keys score (vector data)



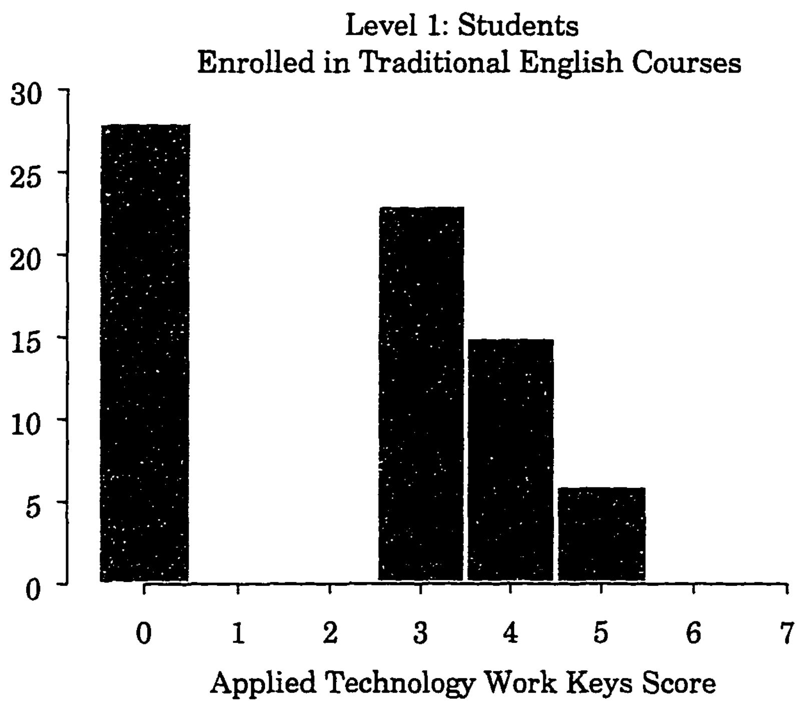


Figure C.29. Histograms comparing Applied Communications versus English students' Applied Technology Work Keys score (vector data)

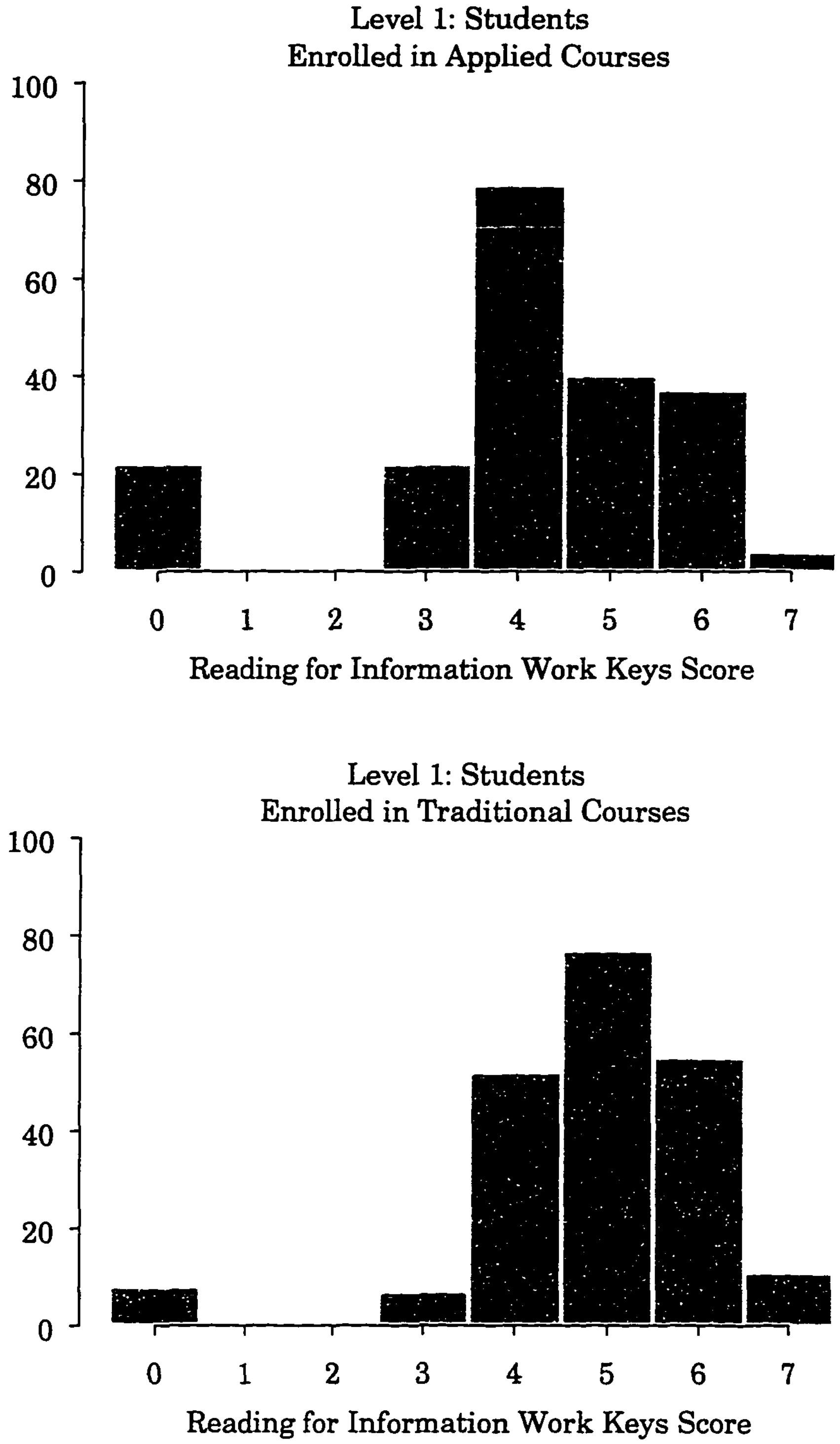


Figure C.30. Histograms comparing "applied" versus "traditional" students' Reading for Information Work Keys score (vector data)

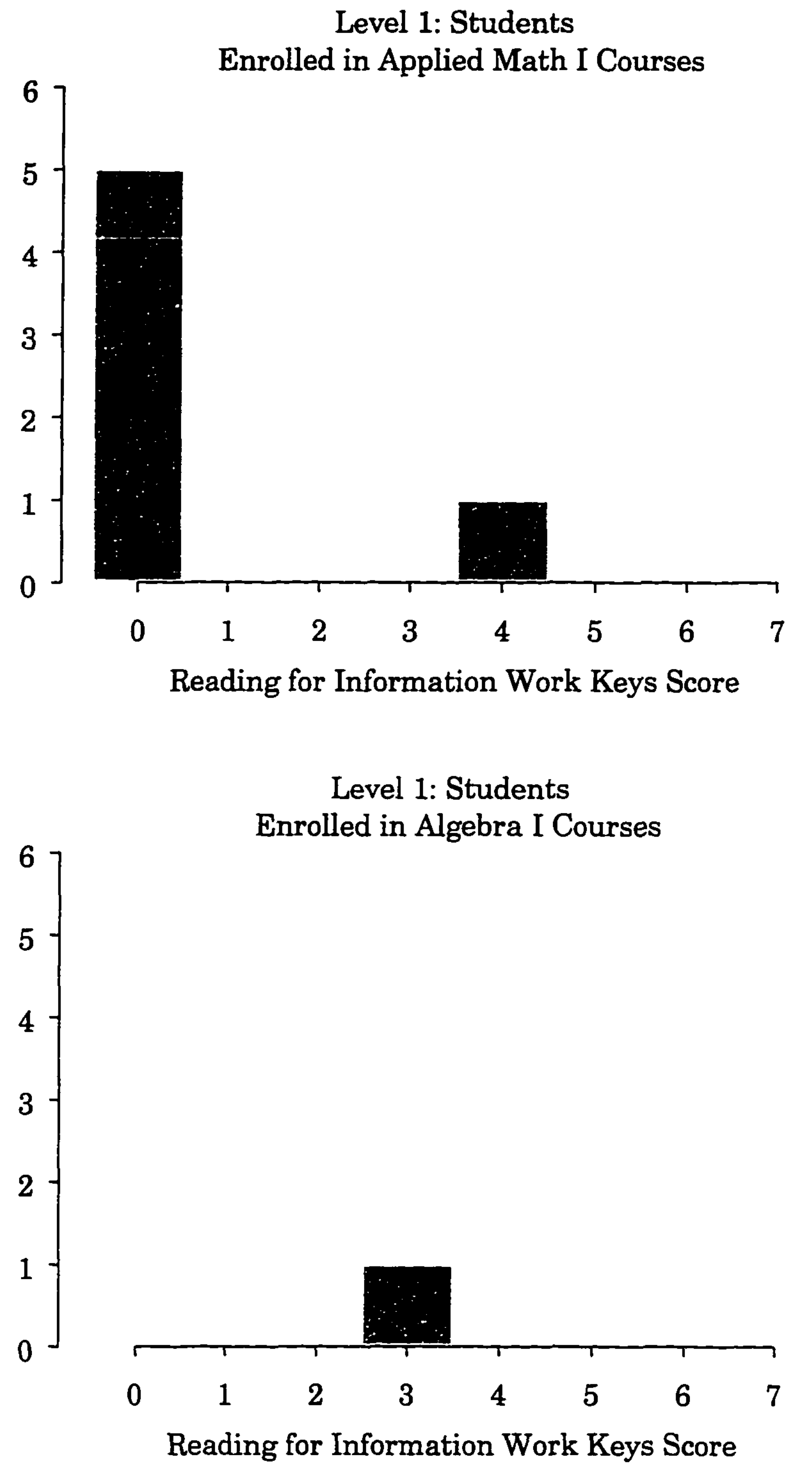
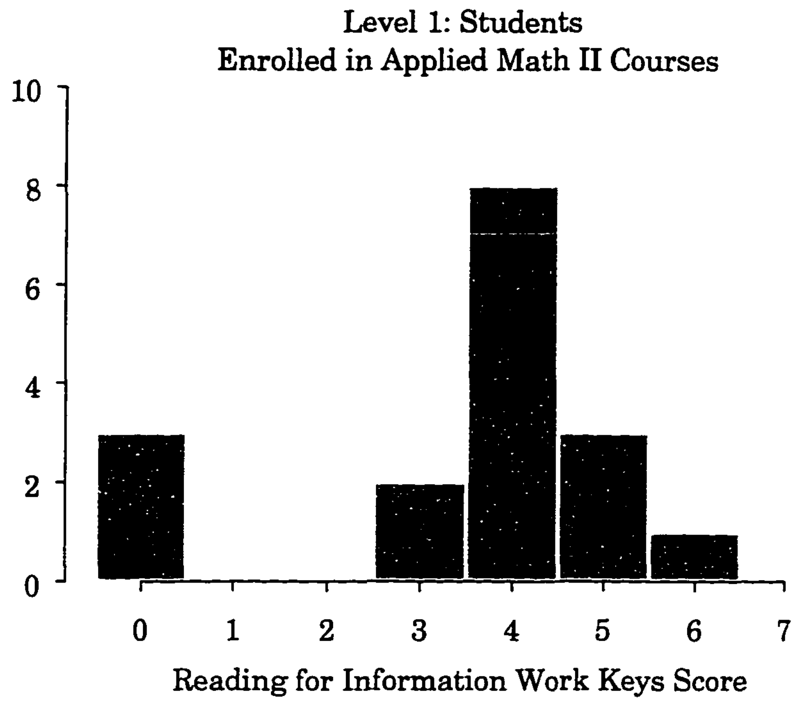


Figure C.31. Histograms comparing Applied Math I versus Algebra I students' Reading for Information Work Keys score (vector data)



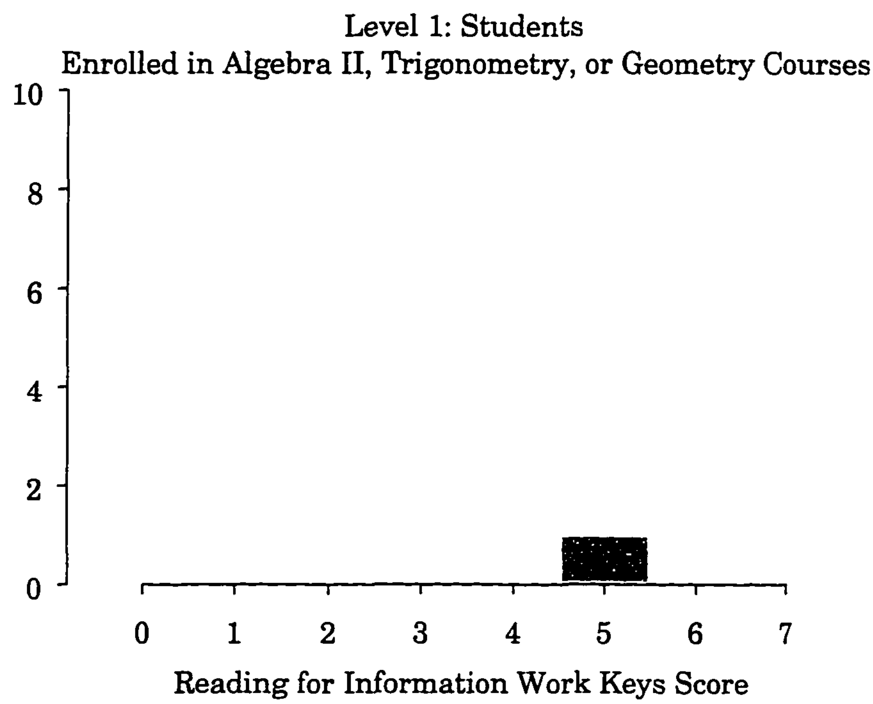


Figure C.32. Histograms comparing Applied Math II versus Traditional Math II students' Reading for Information Work Keys score (vector data)

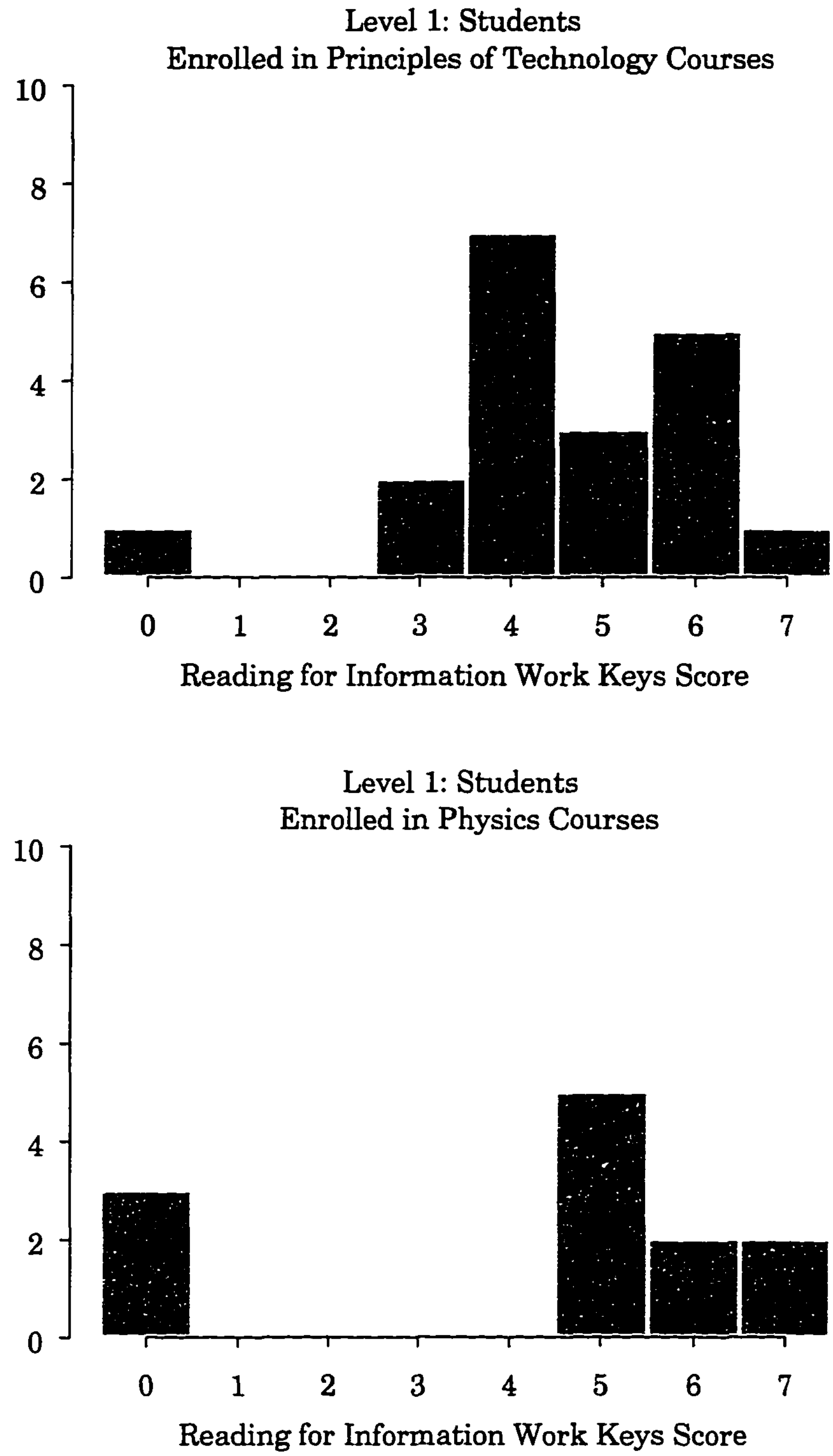


Figure C.33. Histograms comparing Principles of Technology versus Physics students' Reading for Information score (vector data)

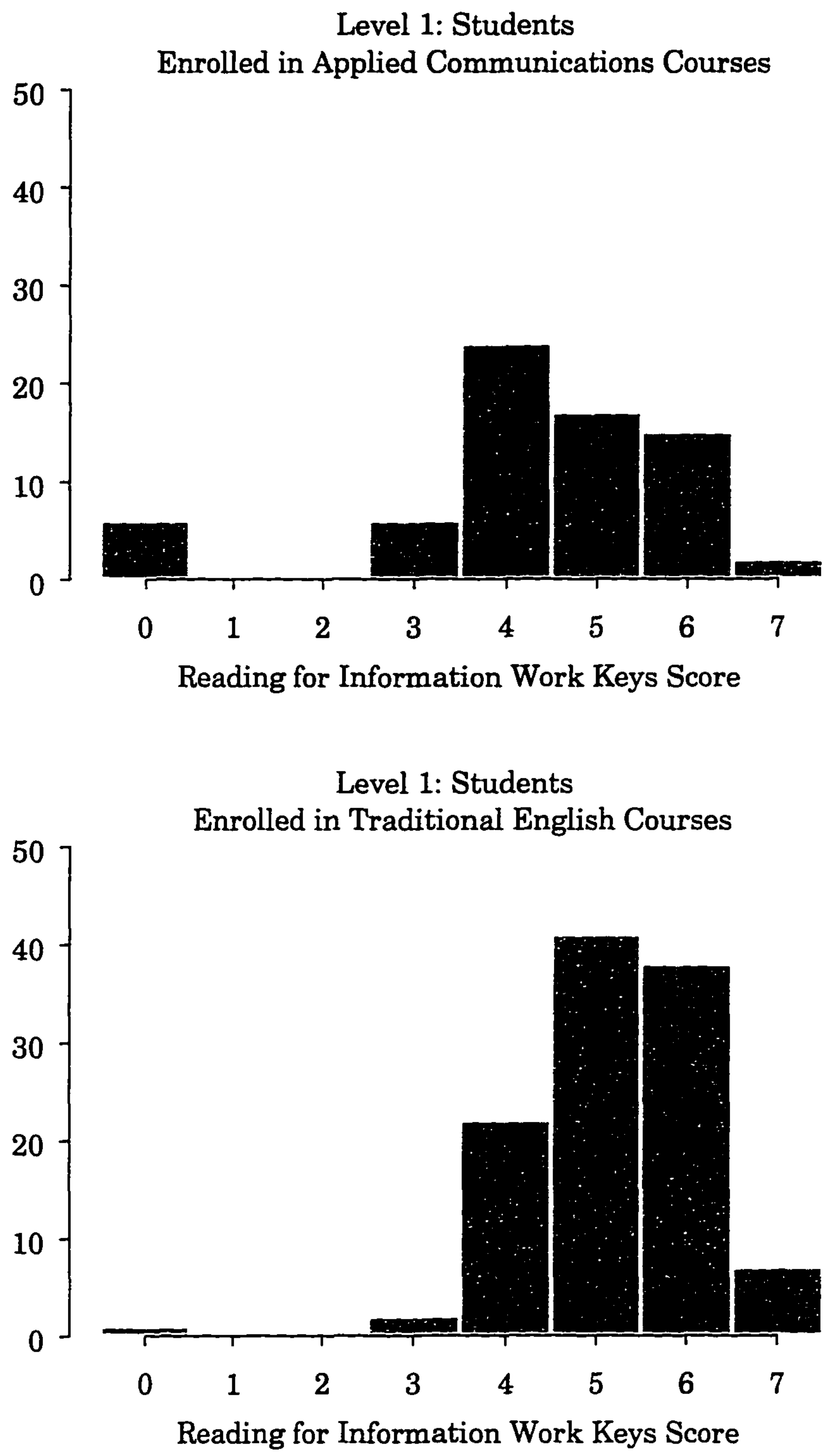
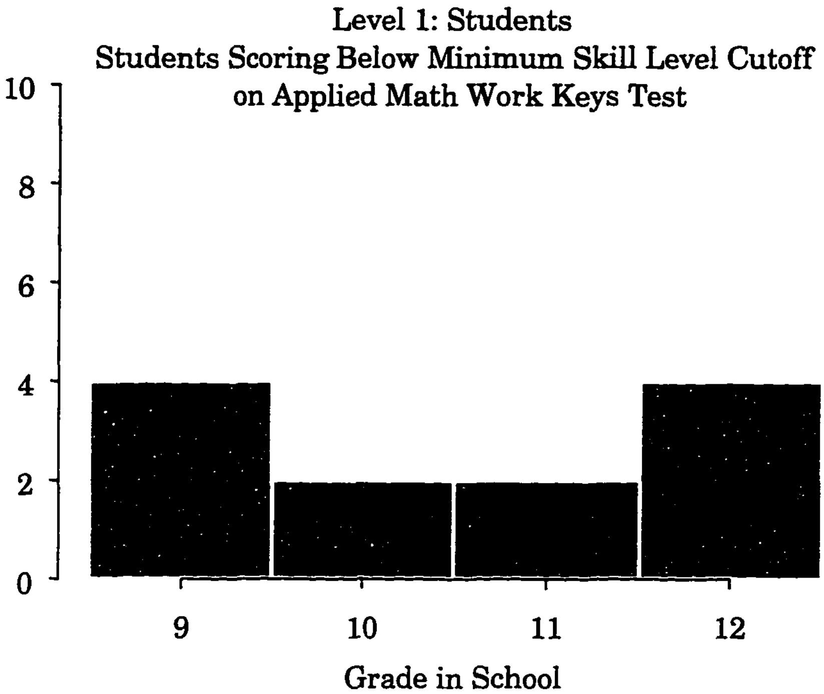


Figure C.34. Histograms comparing Applied Communications versus English students' Reading for Information score (vector data)



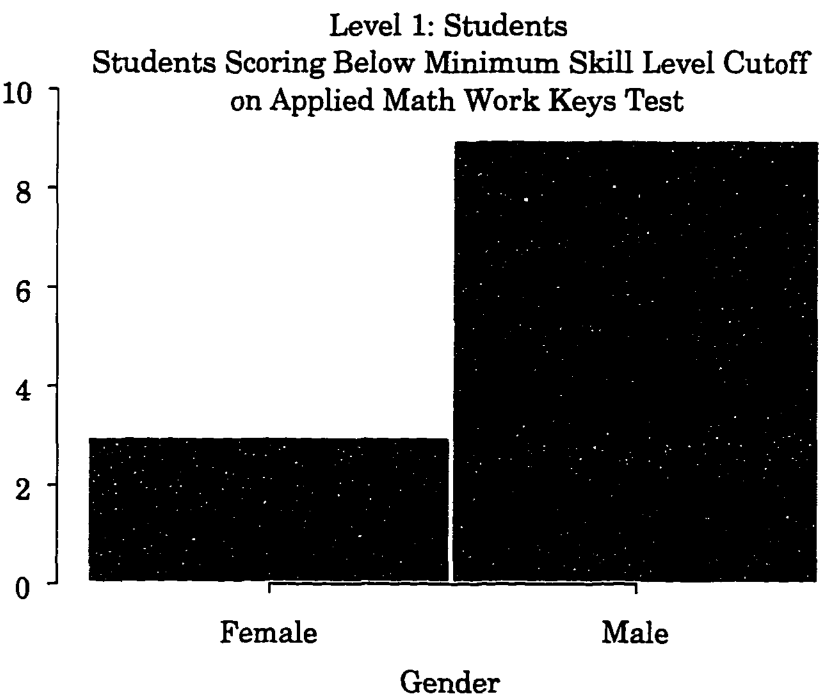
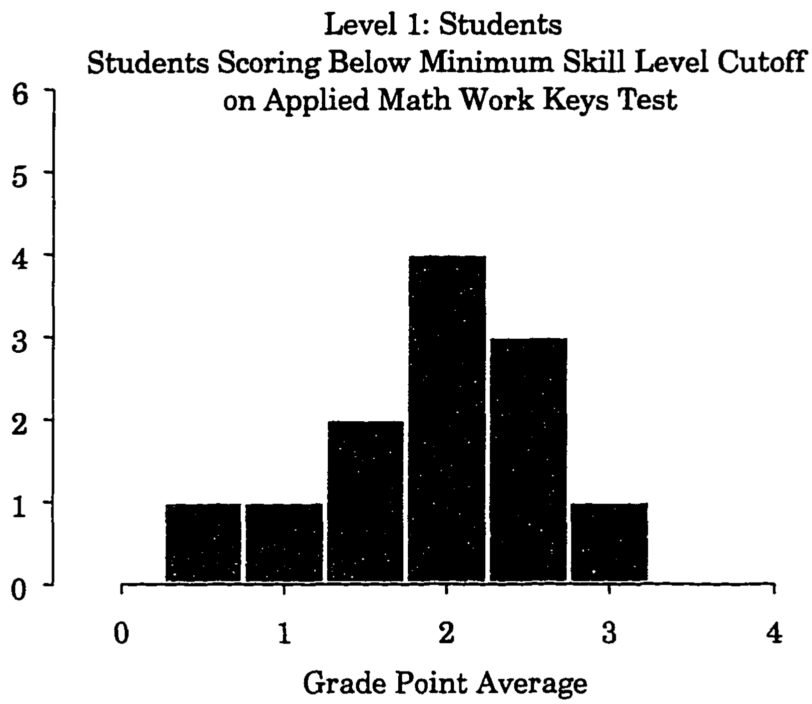


Figure C.35. Grade and Gender histograms of students scoring below the minimum skill level (3) on AM Work Keys test (full matrix data)



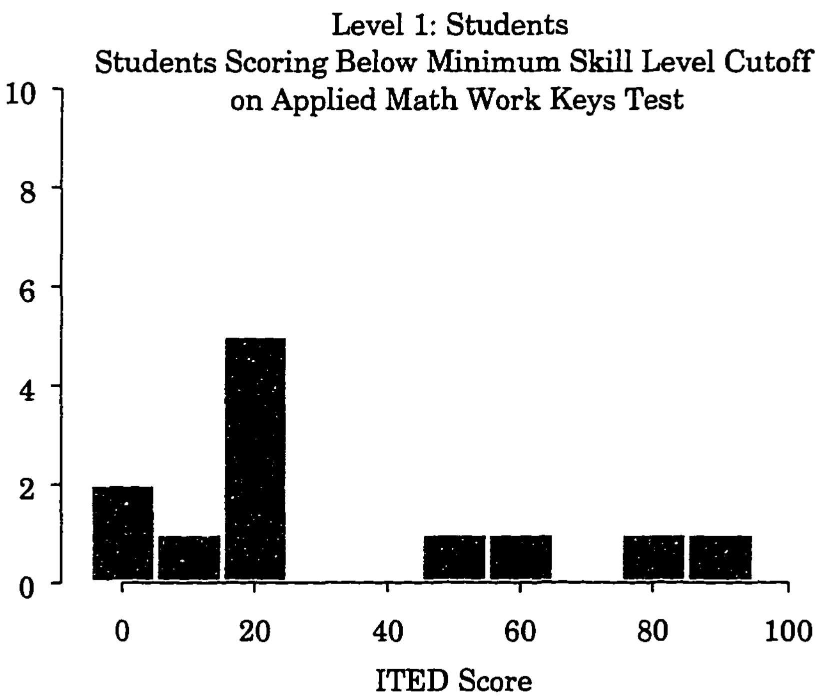
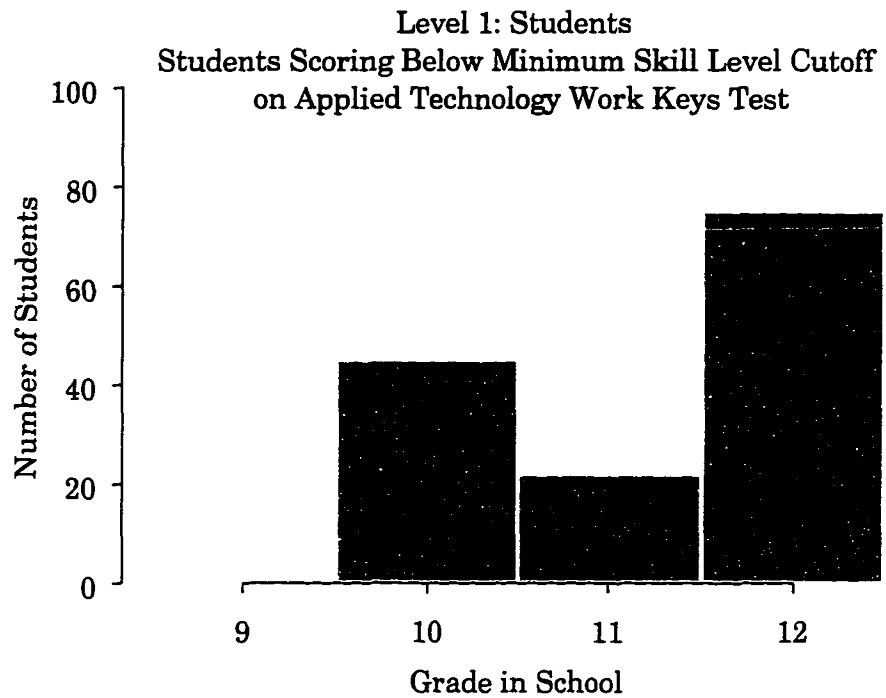


Figure C.36. GPA and ITED histograms of students scoring below the minimum skill level (3) on AM Work Keys test (full matrix data)



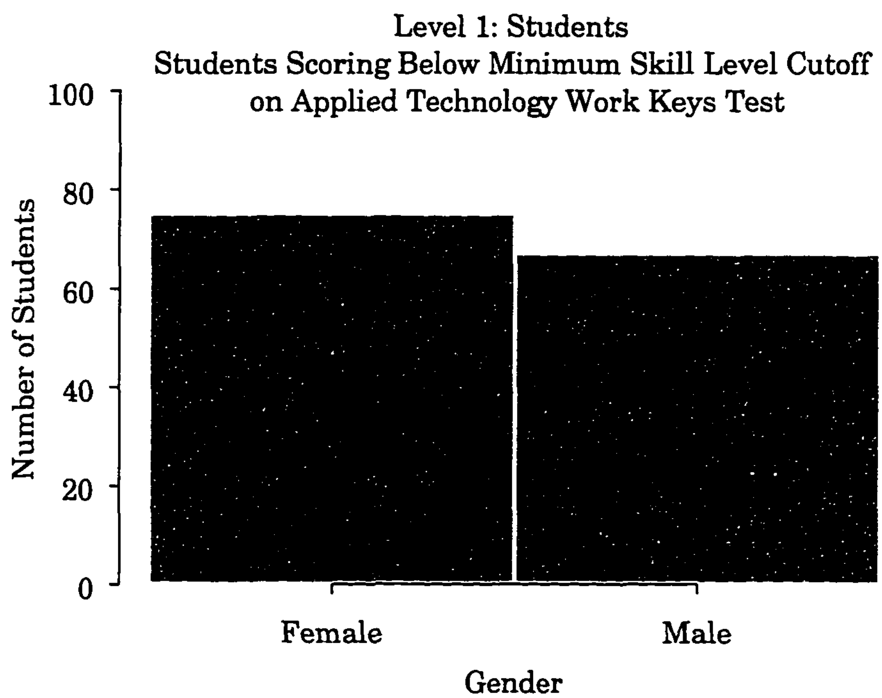
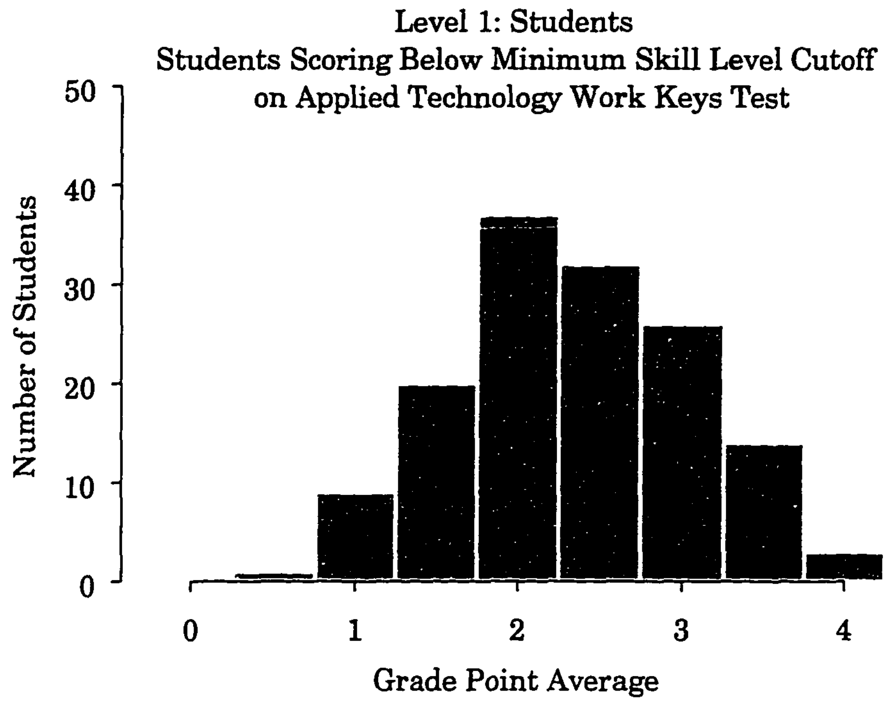


Figure C.37. Grade and Gender histograms of students scoring below the minimum skill level (3) on AT Work Keys test (full matrix data)



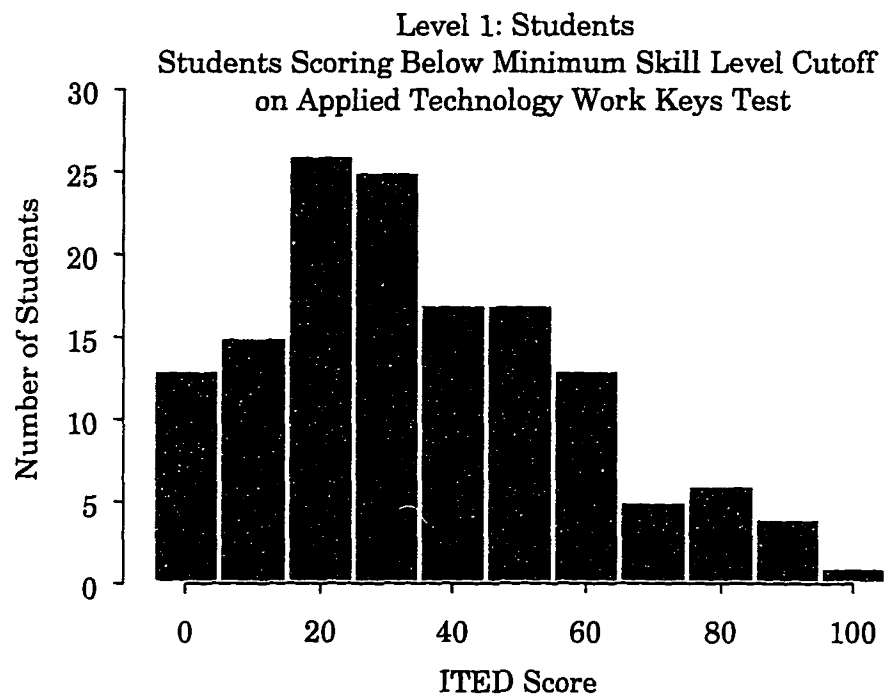
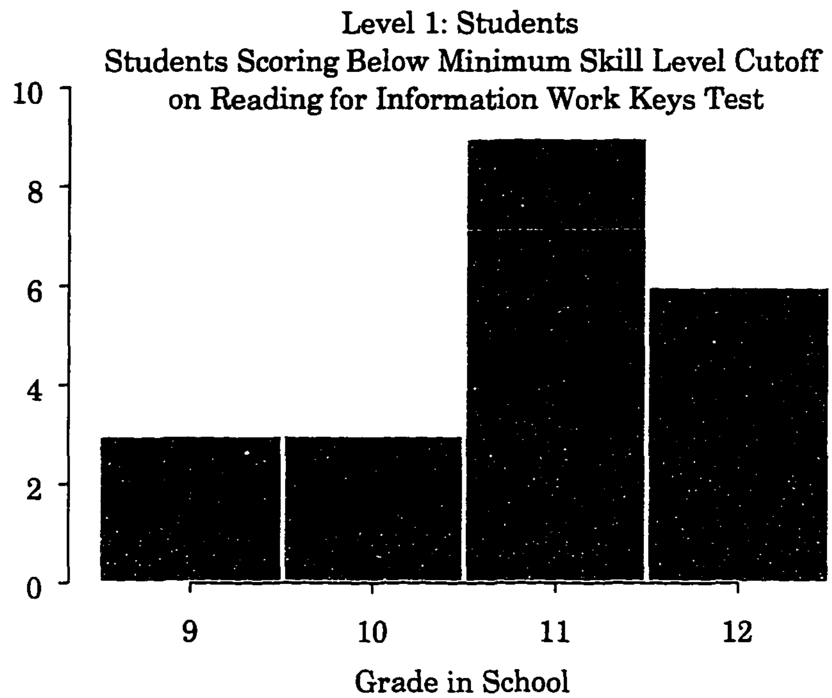


Figure C.38. GPA and ITED histograms of students scoring below the minimum skill level (3) on AT Work Keys test (full matrix data)



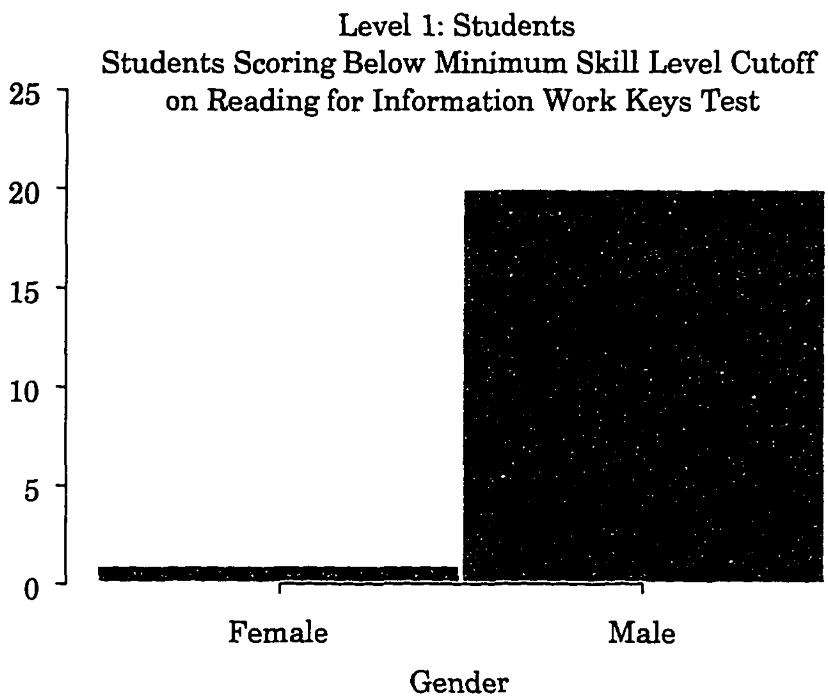
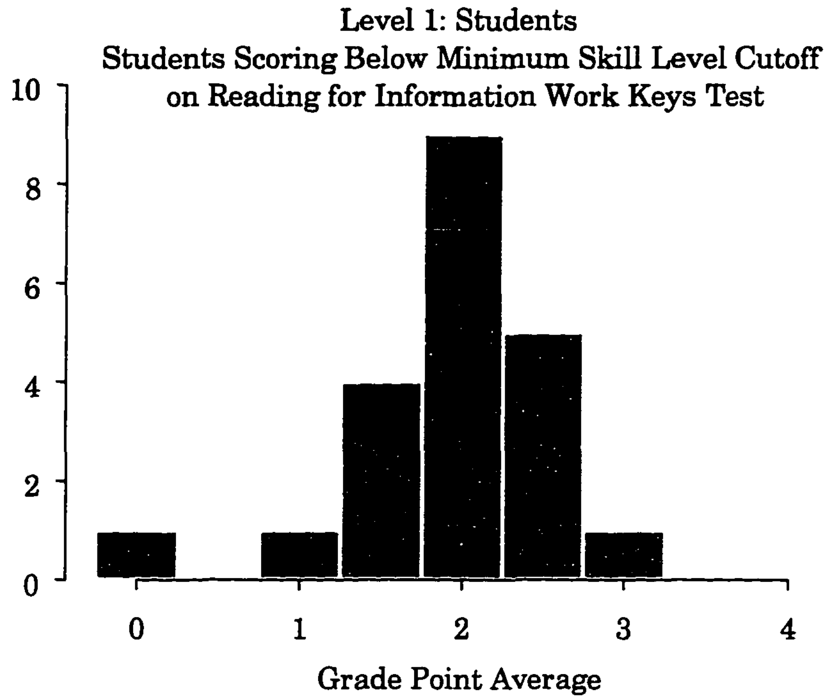


Figure C.39. Grade and Gender histograms of students scoring below the minimum skill level (3) on RFI Work Keys test (full matrix data)



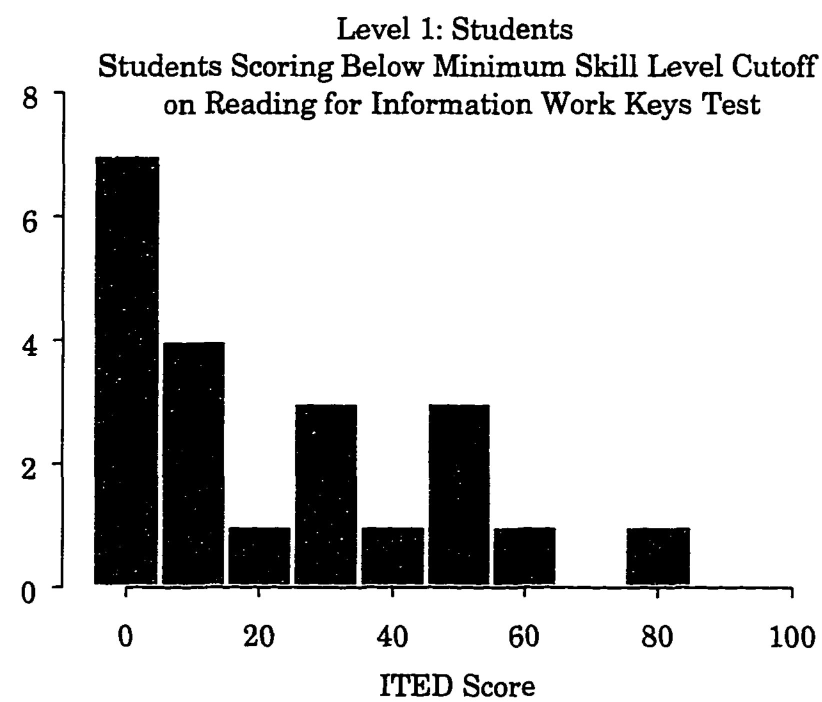
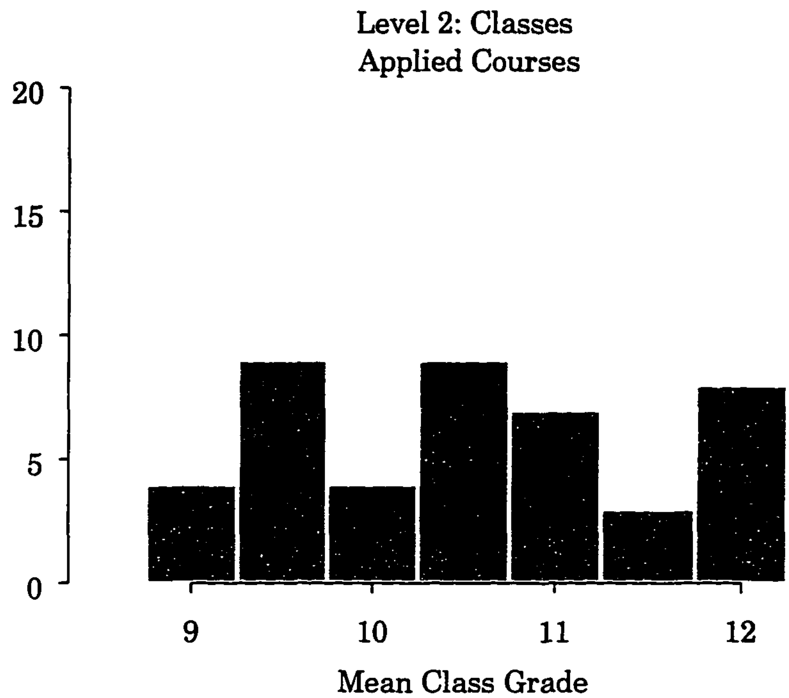


Figure C.40. GPA and ITED histograms of students scoring below the minimum skill level (3) on RFI Work Keys test (full matrix data)



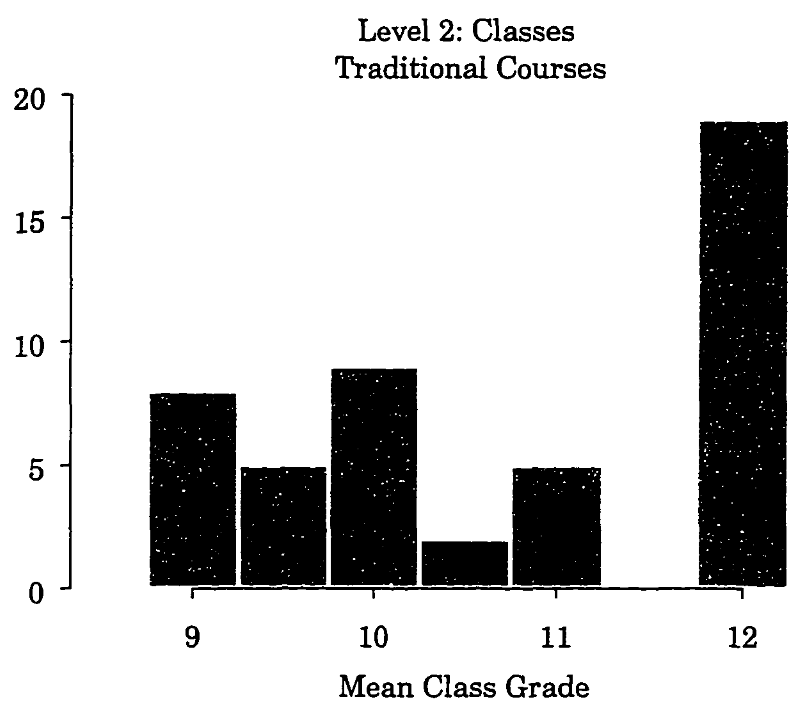
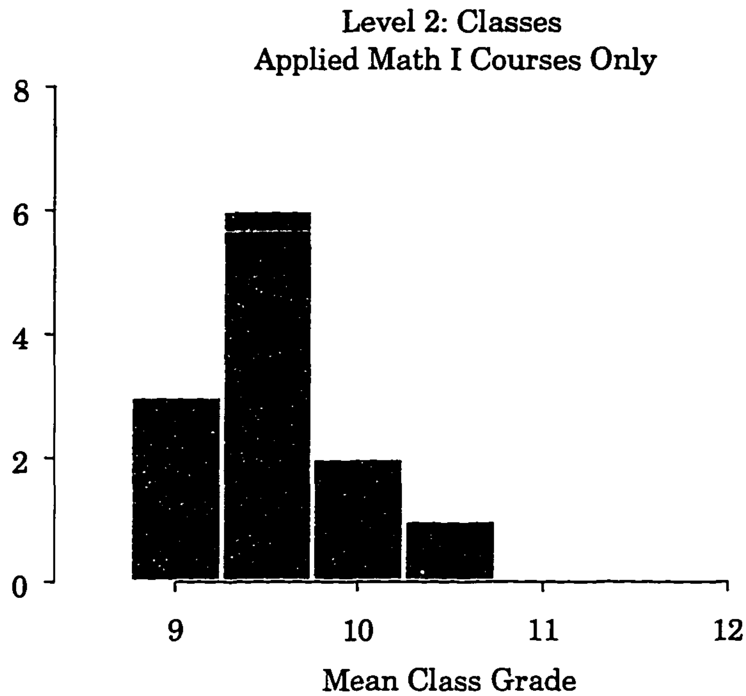


Figure C.41. Histograms comparing mean class grade of applied classes versus traditional classes (vector data)



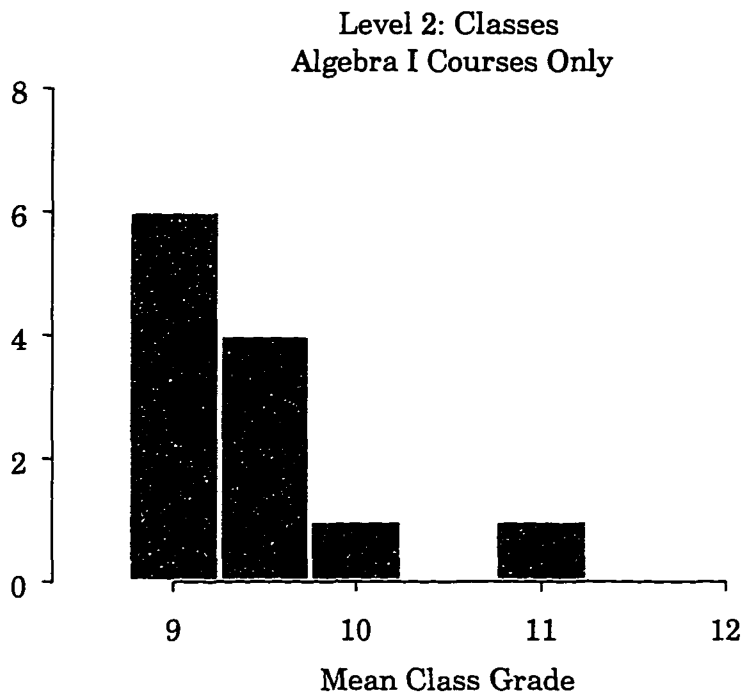
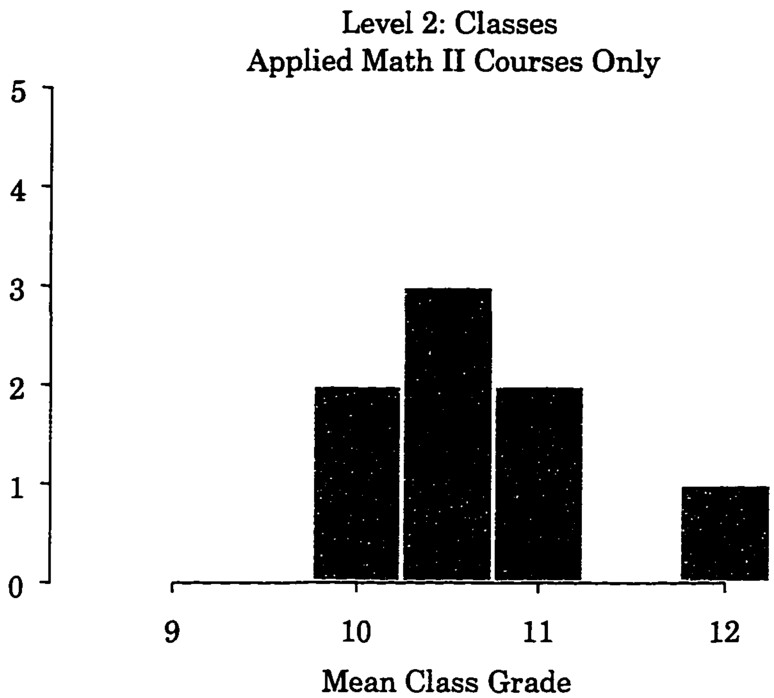


Figure C.42. Histograms comparing mean class grade of Applied Math I versus Algebra I classes (vector data)



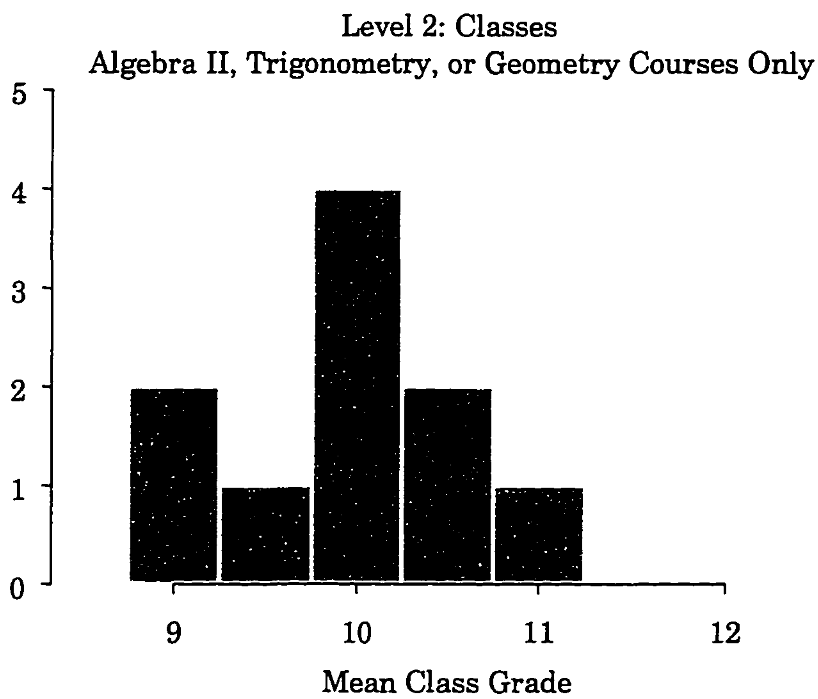


Figure C.43. Histograms comparing mean class grade of Applied Math II versus Traditional Math II classes (vector data)

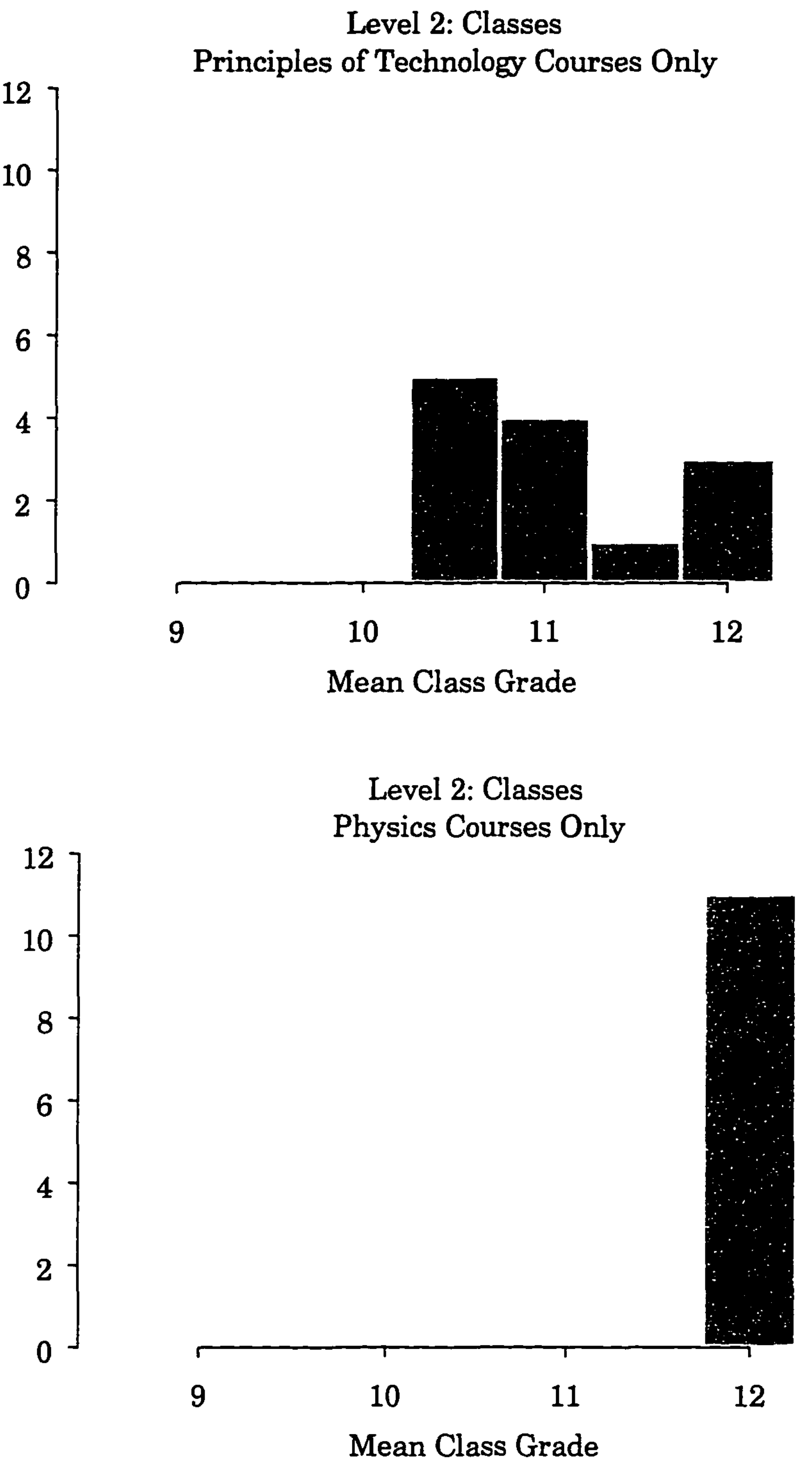
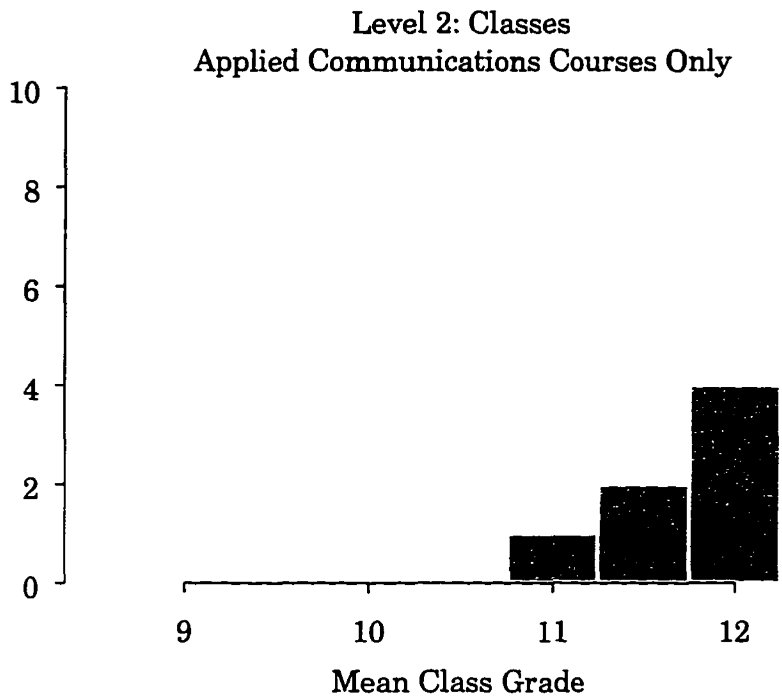


Figure C.44. Histograms comparing mean class grade of Principles of Technology versus Physics classes (vector data)



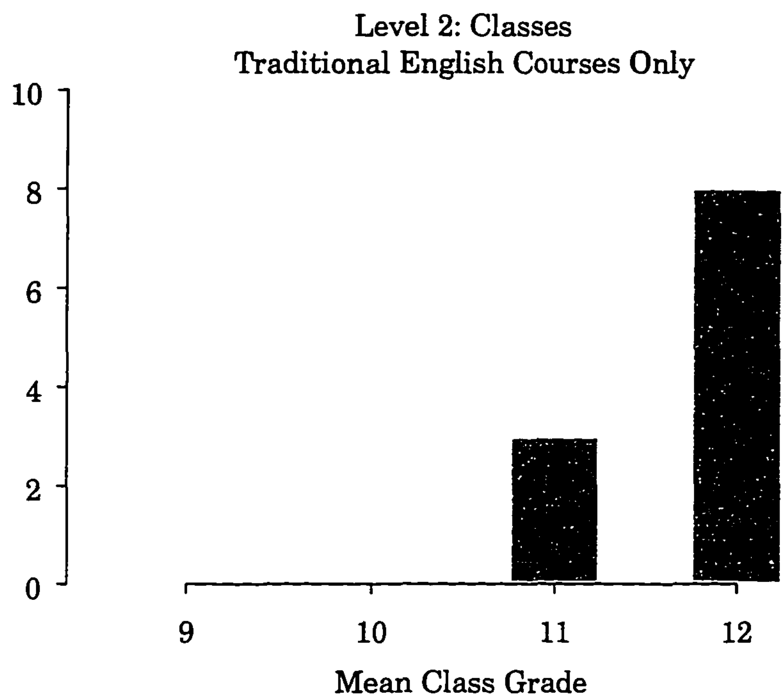
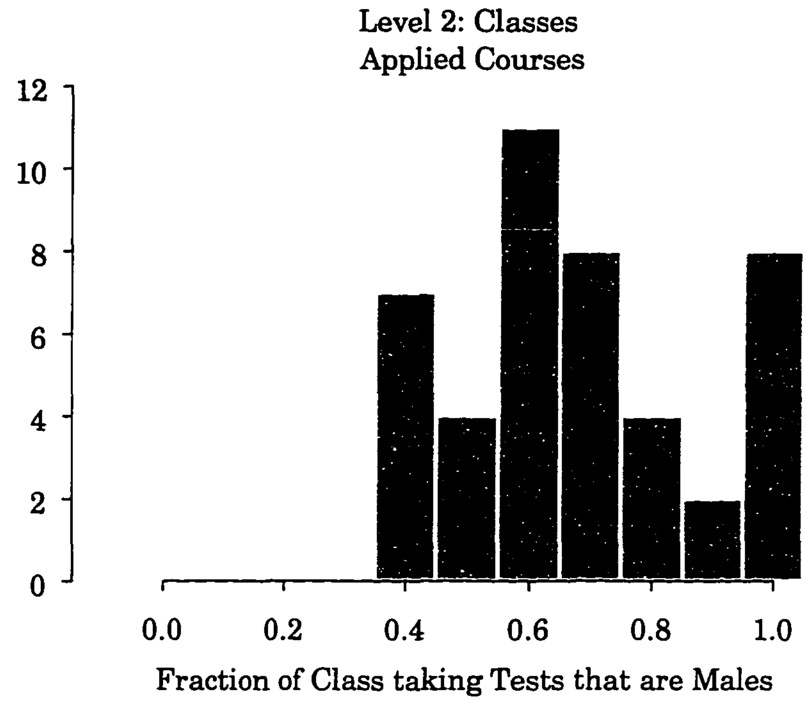


Figure C.45. Histograms comparing mean class grade of Applied Communications versus English classes (vector data)



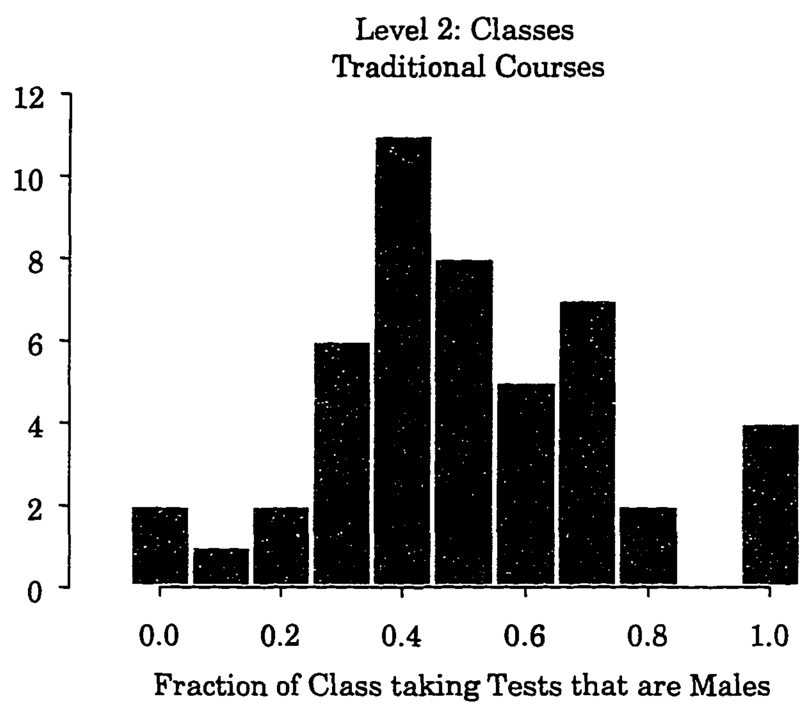
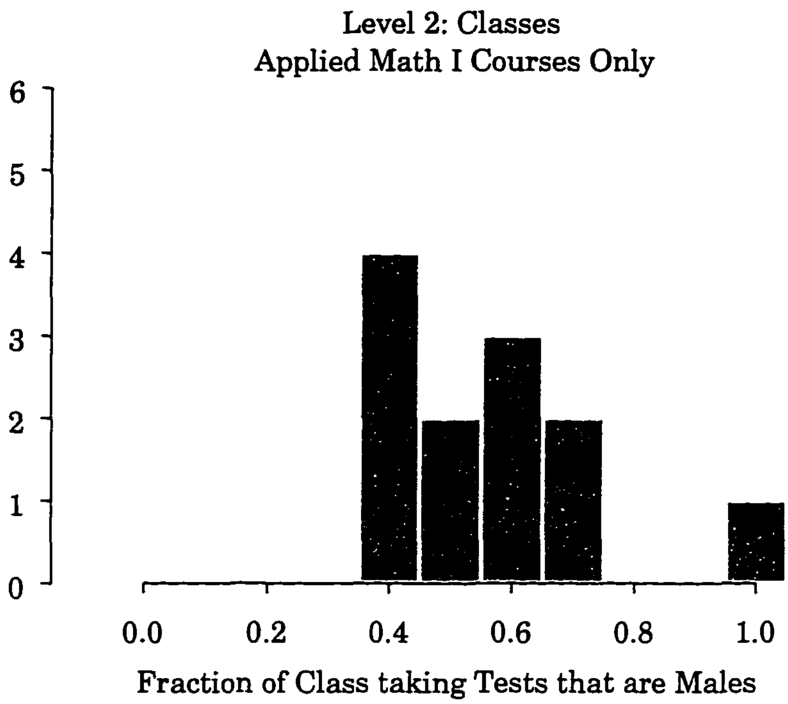


Figure C.46. Histograms comparing fraction of class that is male of applied classes versus traditional classes (vector data)



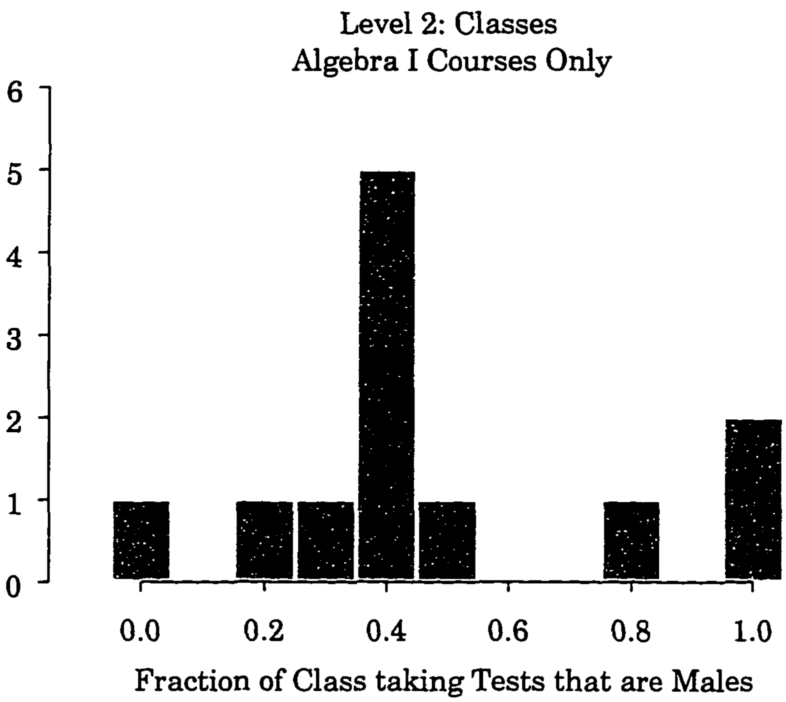


Figure C.47. Histograms comparing fraction of class that is male of Applied Math I versus Algebra I classes (vector data)

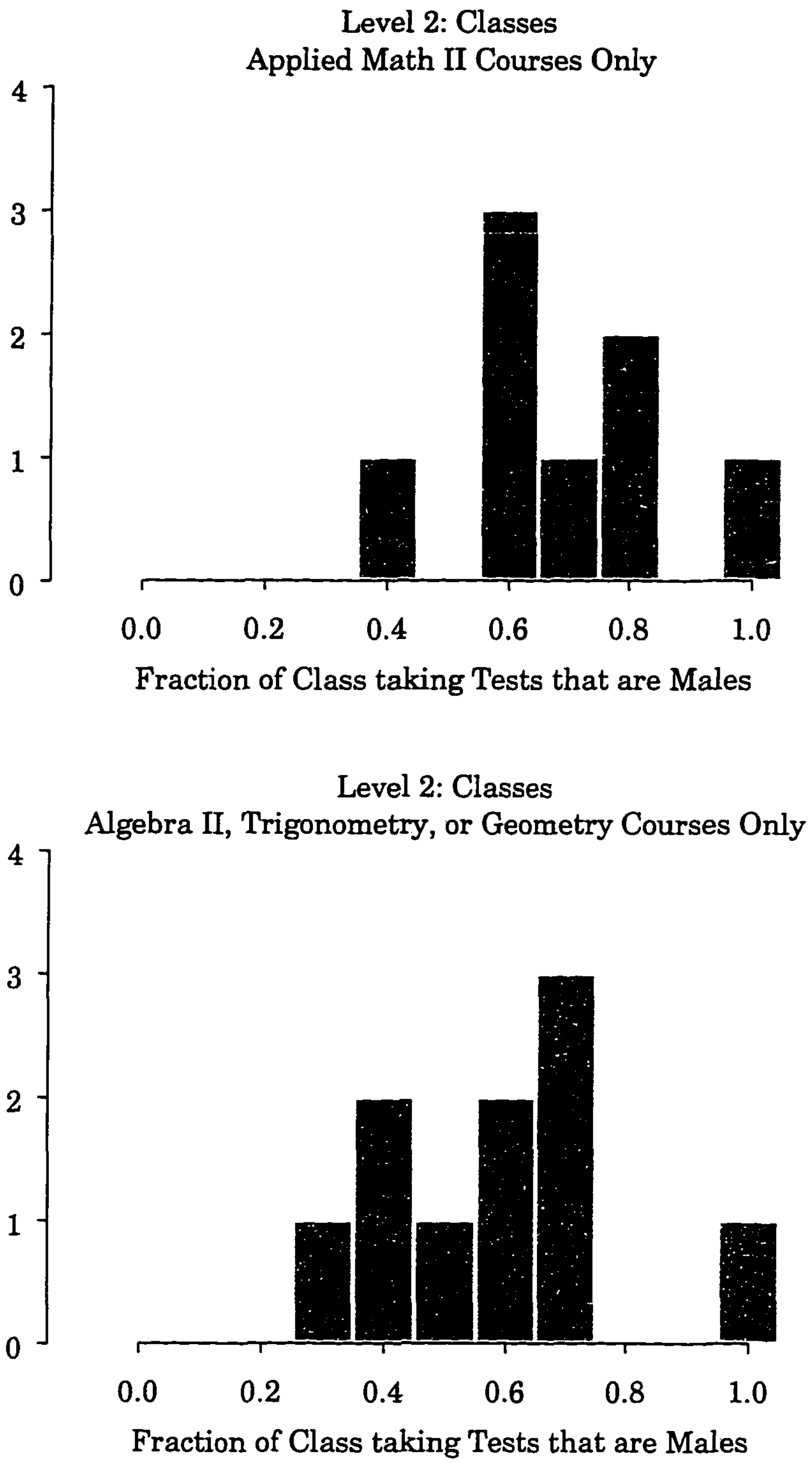


Figure C.48. Histograms comparing fraction of class that is male of Applied Math II versus Traditional Math II classes (vector data)

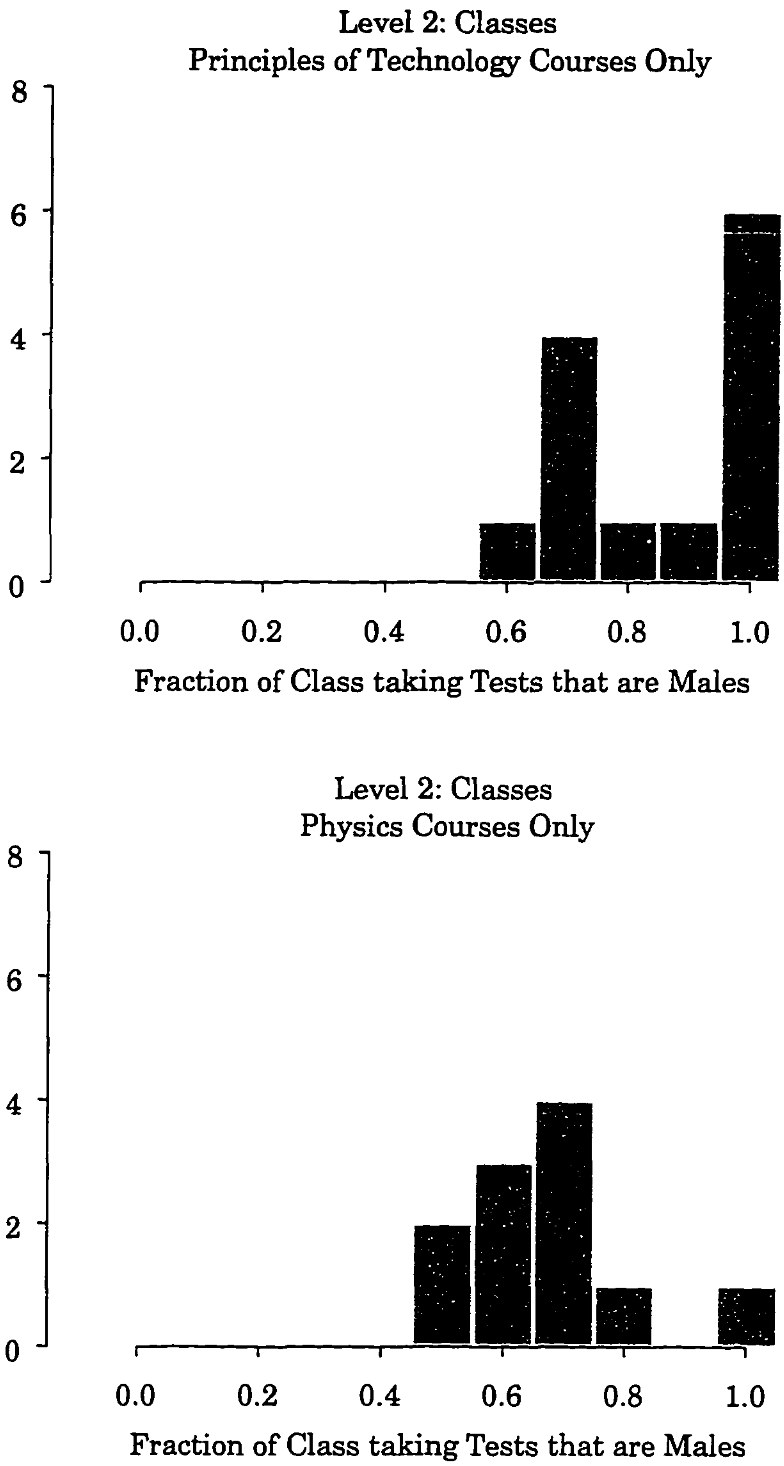


Figure C.49. Histograms comparing fraction of class that is male of Principles of Technology versus Physics classes (vector data)

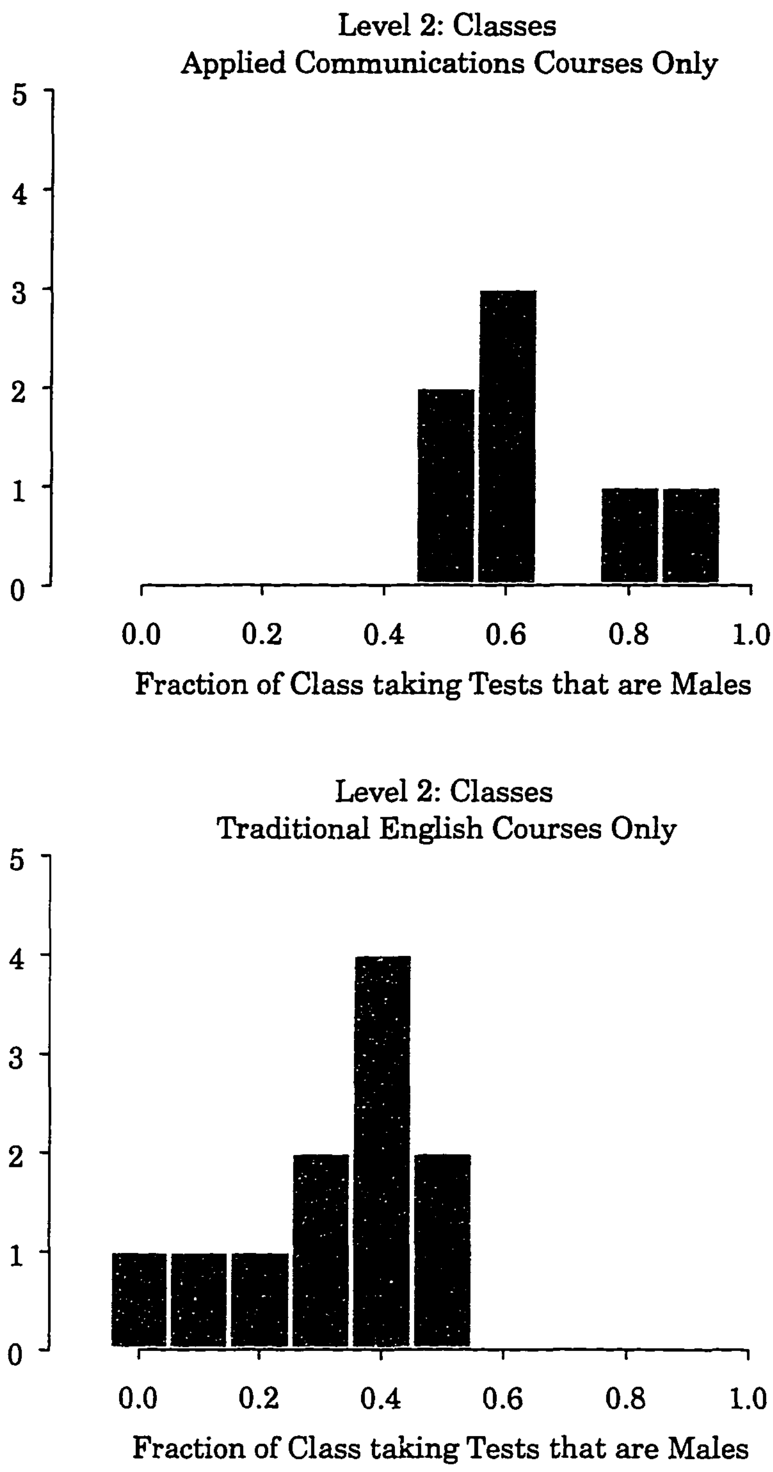


Figure C.50. Histograms comparing fraction of class that is male of Applied Communications versus Traditional English classes (vector data)

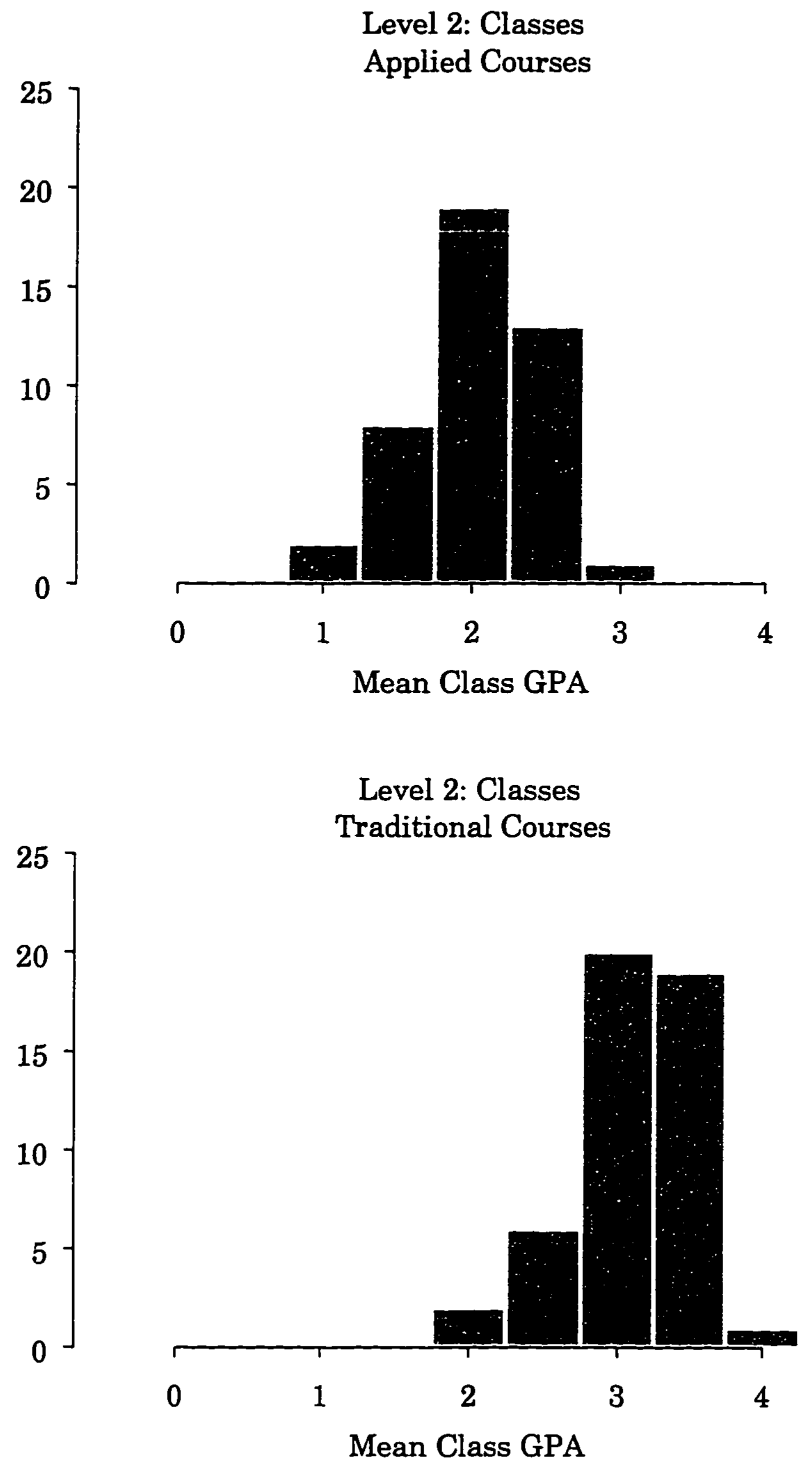
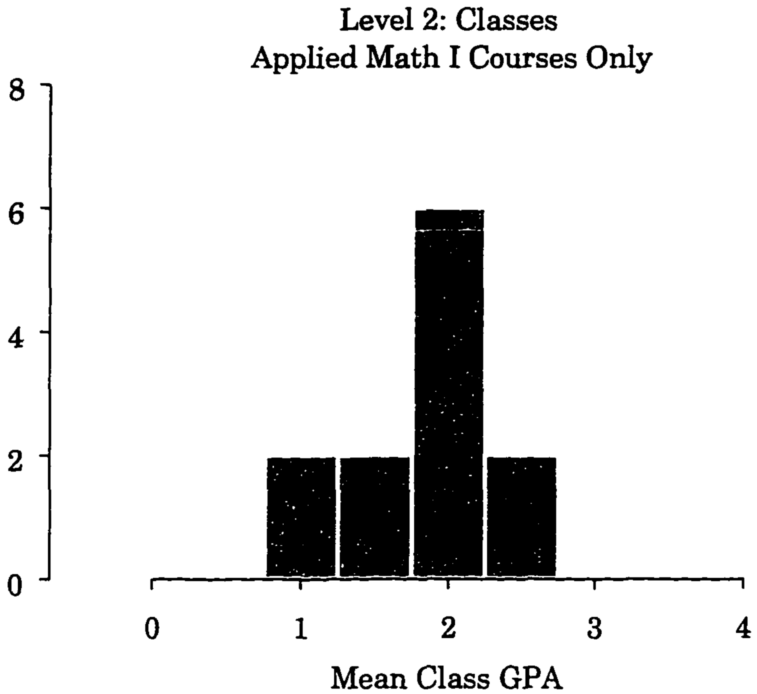


Figure C.51. Histograms comparing mean class GPA of applied versus traditional classes (vector data)



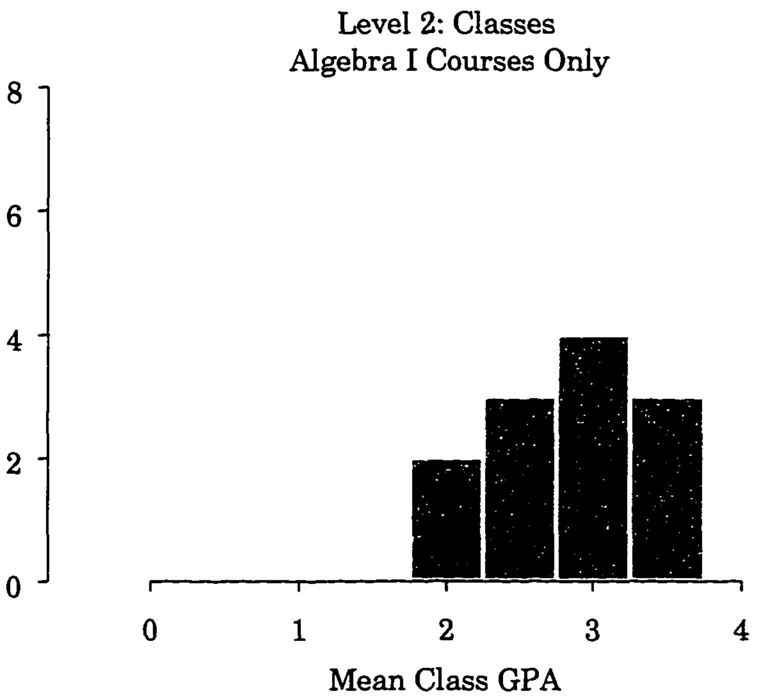


Figure C.52. Histograms comparing mean class GPA of Applied Math I versus Algebra I classes (vector data)

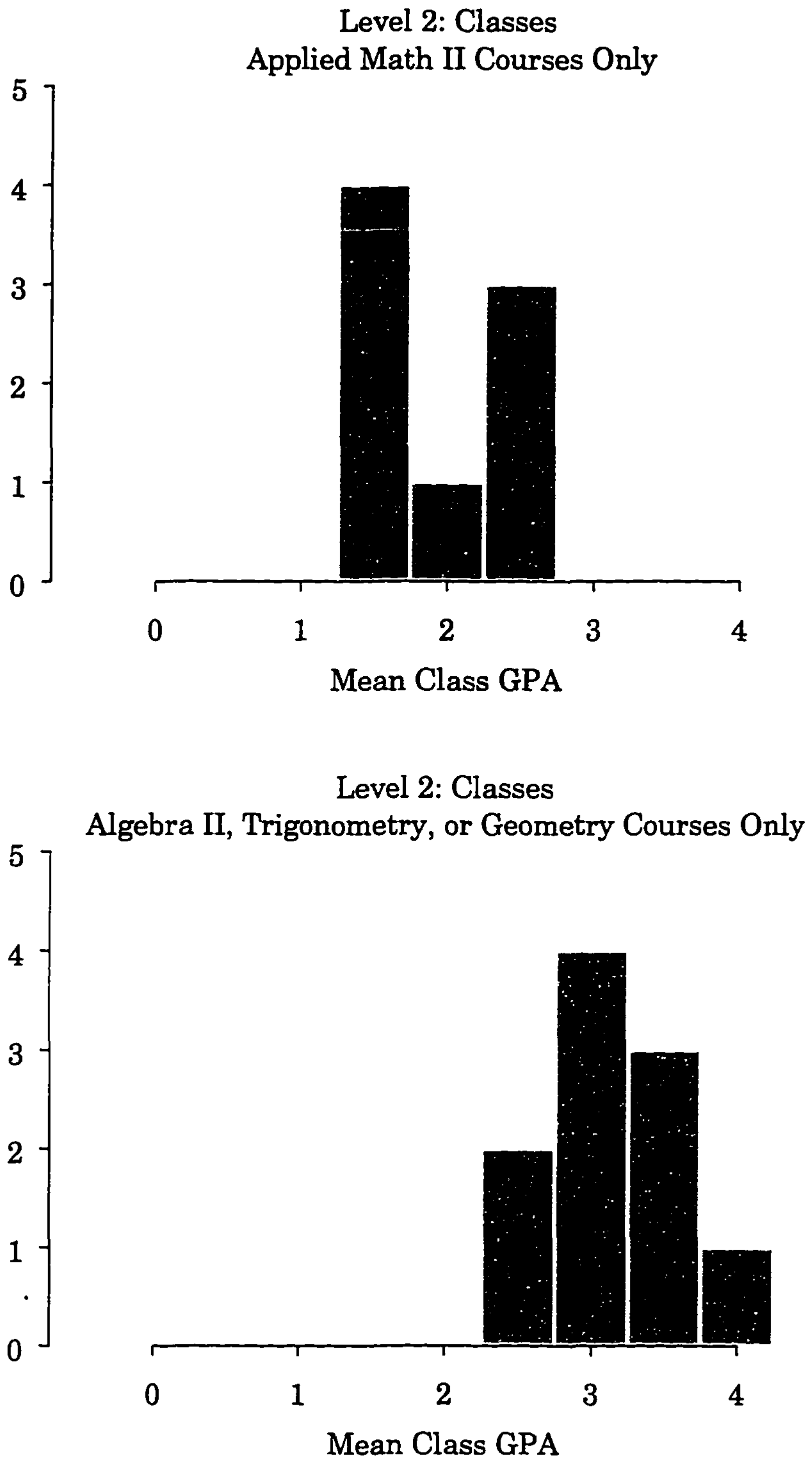


Figure C.53. Histograms comparing mean class GPA of Applied Math II versus Traditional Math II classes (vector data)

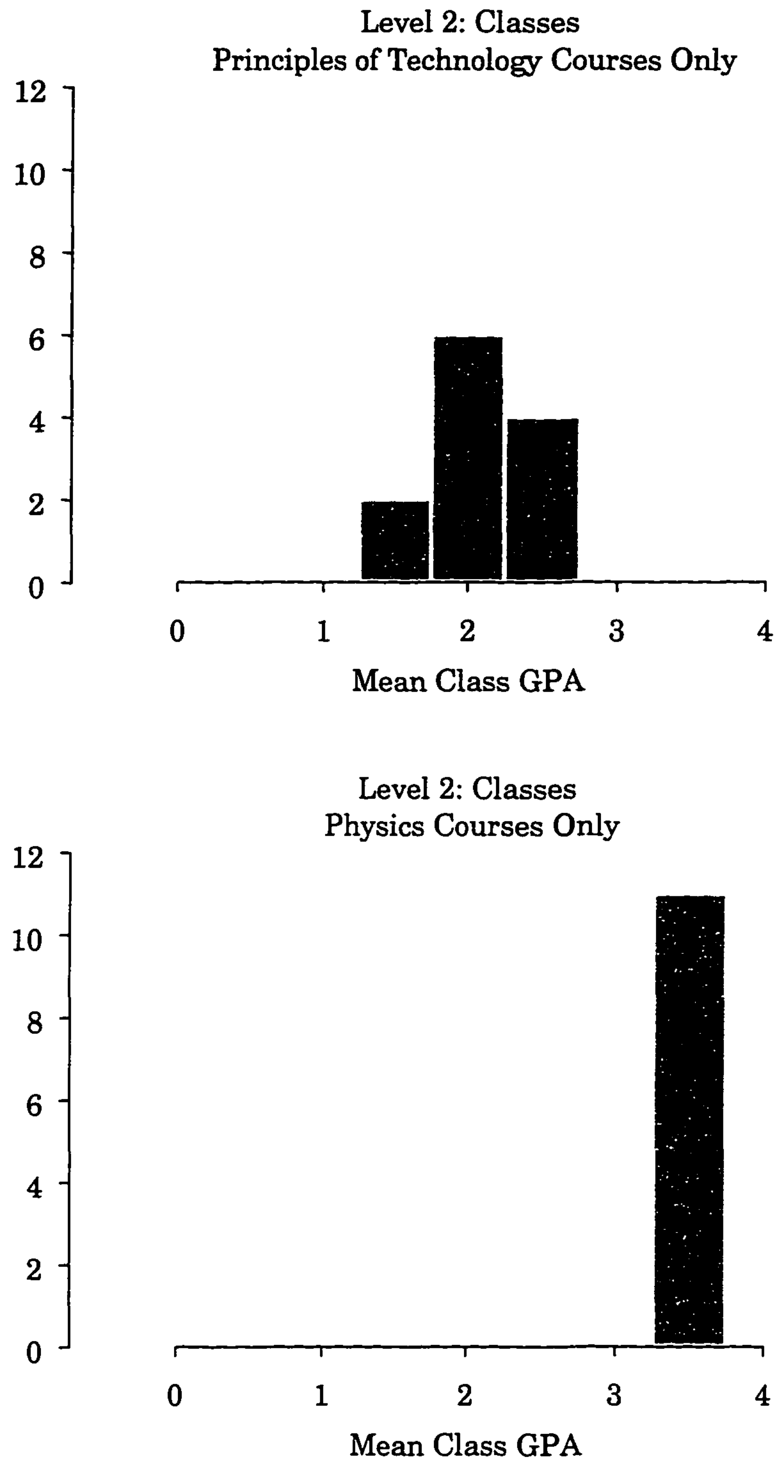
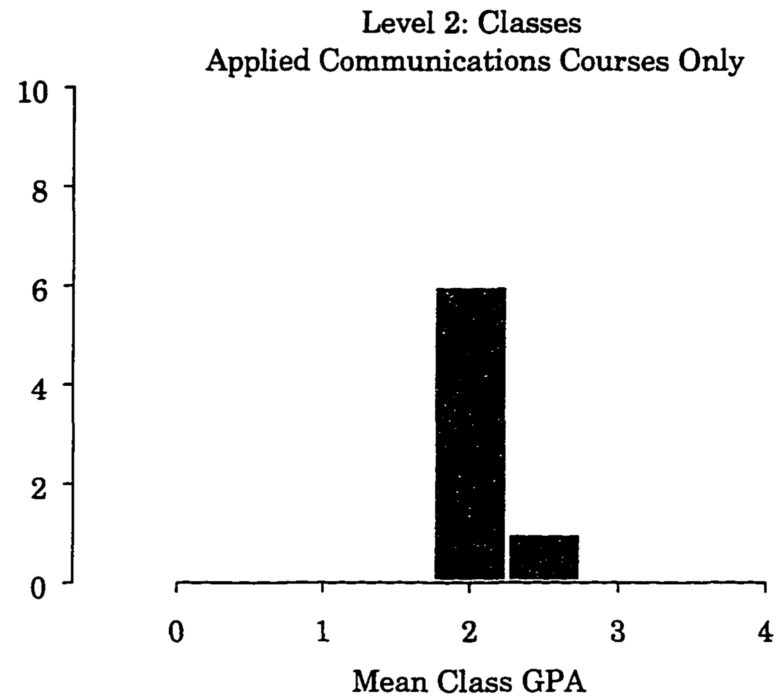


Figure C.54. Histograms comparing mean class GPA of Principles of Technology versus Physics classes (vector data)



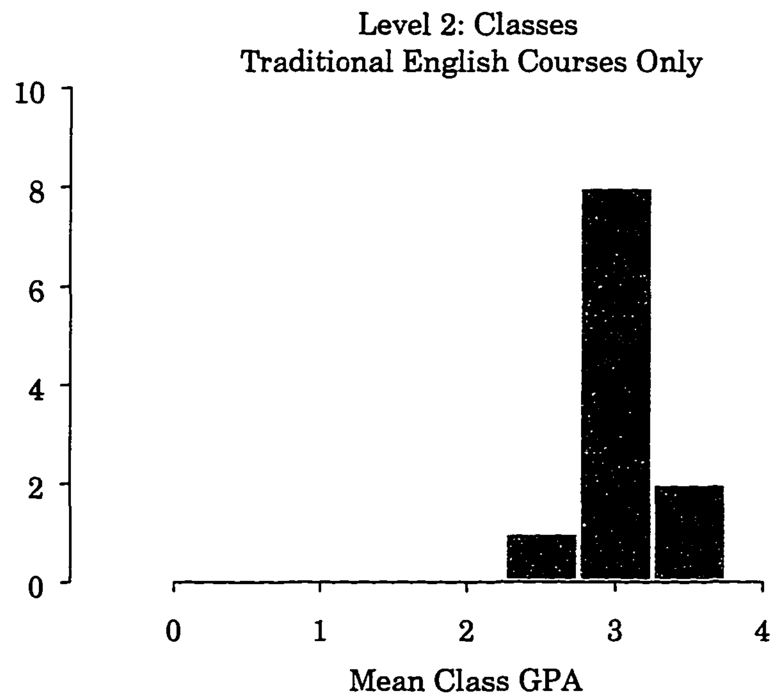
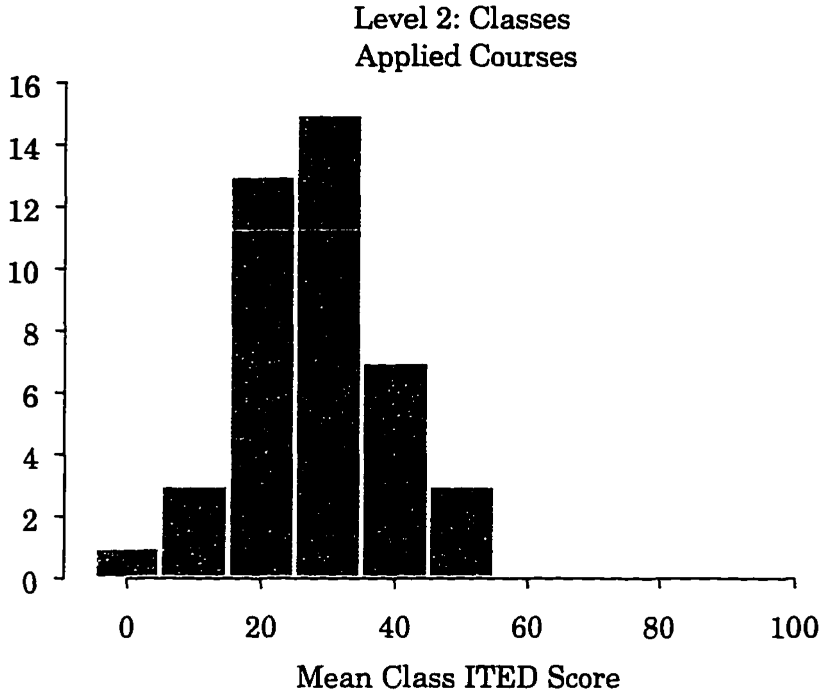


Figure C.55. Histograms comparing mean class GPA of Applied Communications versus Traditional English classes (vector data)



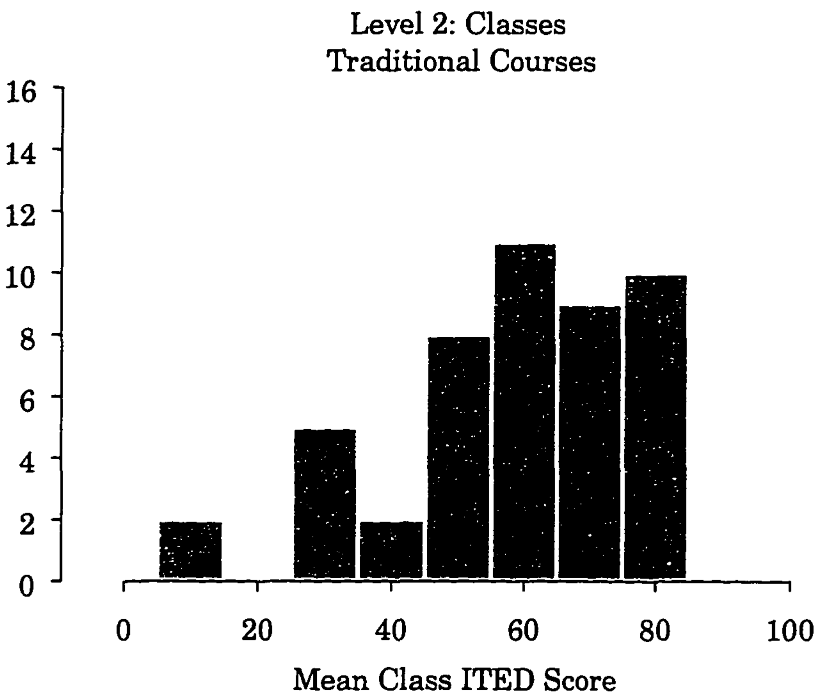


Figure C.56. Histograms comparing mean class ITED score of applied classes versus traditional classes (vector data)

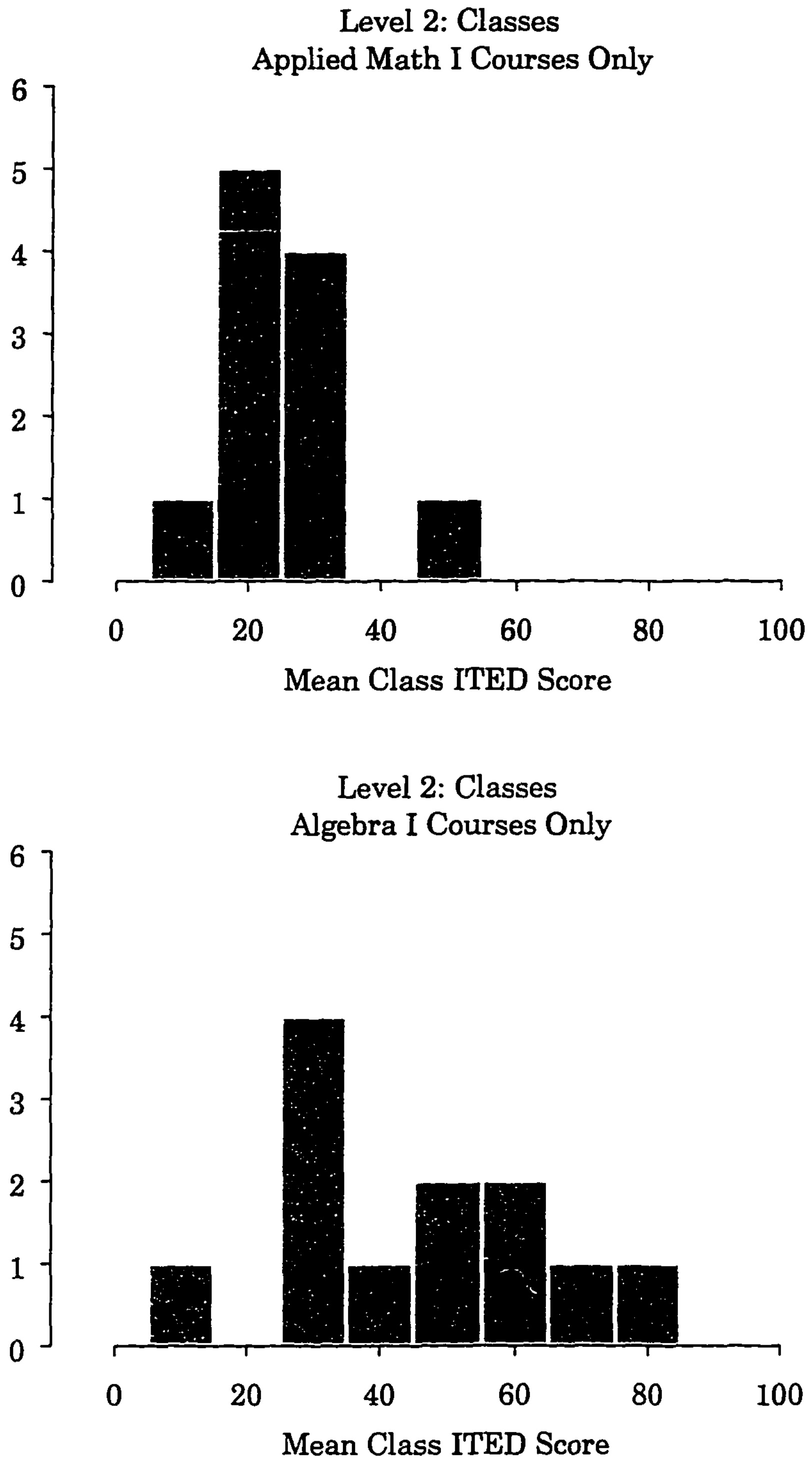
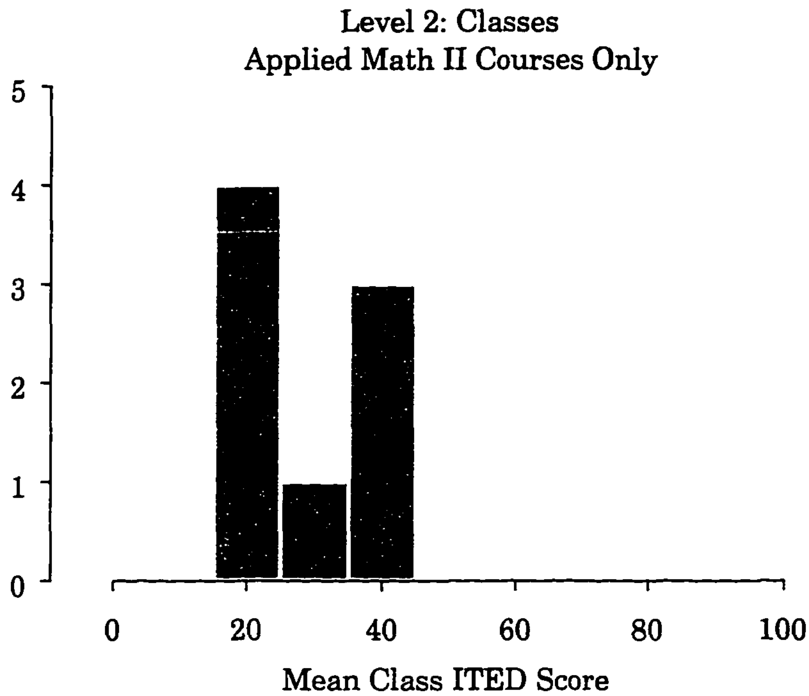


Figure C.57. Histograms comparing mean class ITED score of Applied Math I versus Algebra I classes (vector data)



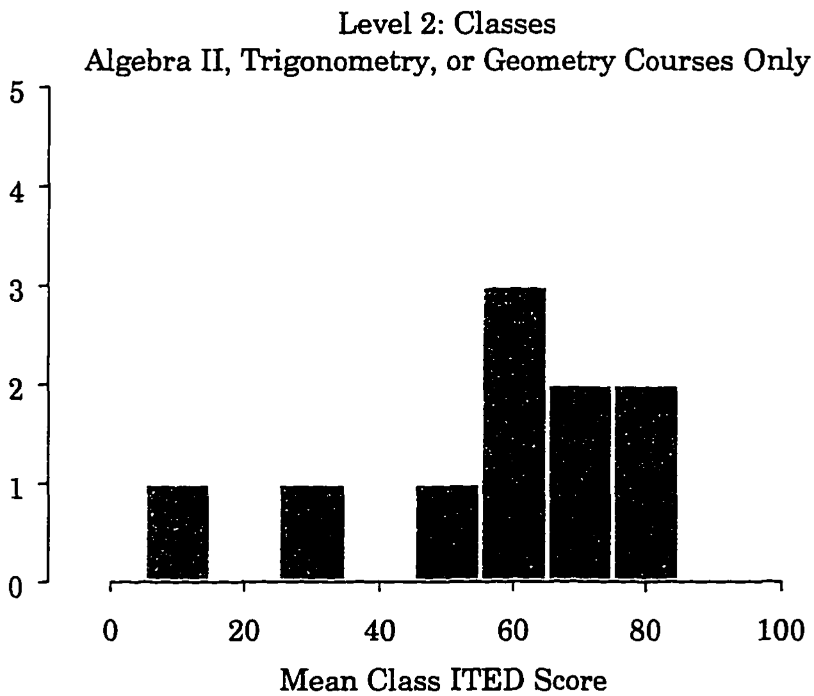
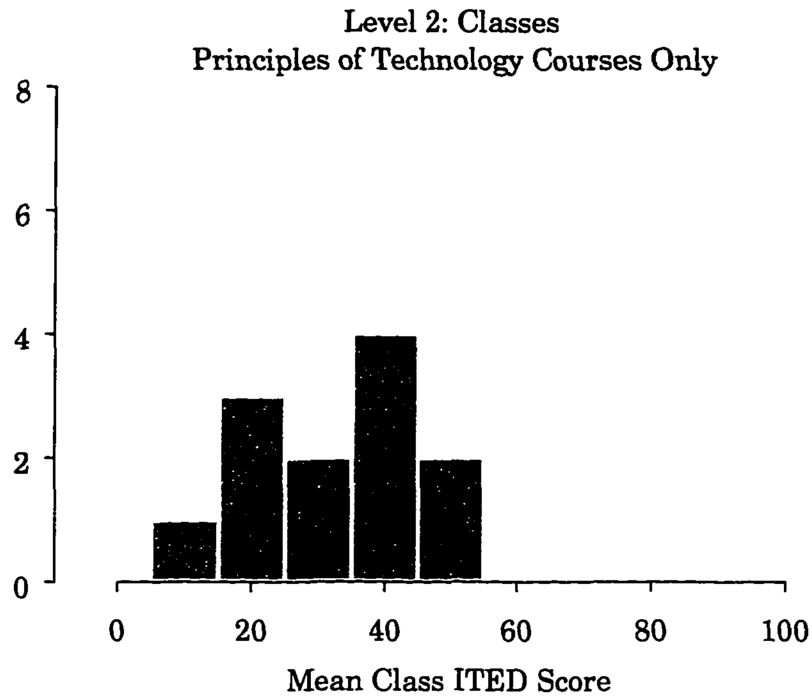


Figure C.58. Histograms comparing mean class ITED score of Applied Math II versus Traditional Math II classes (vector data)



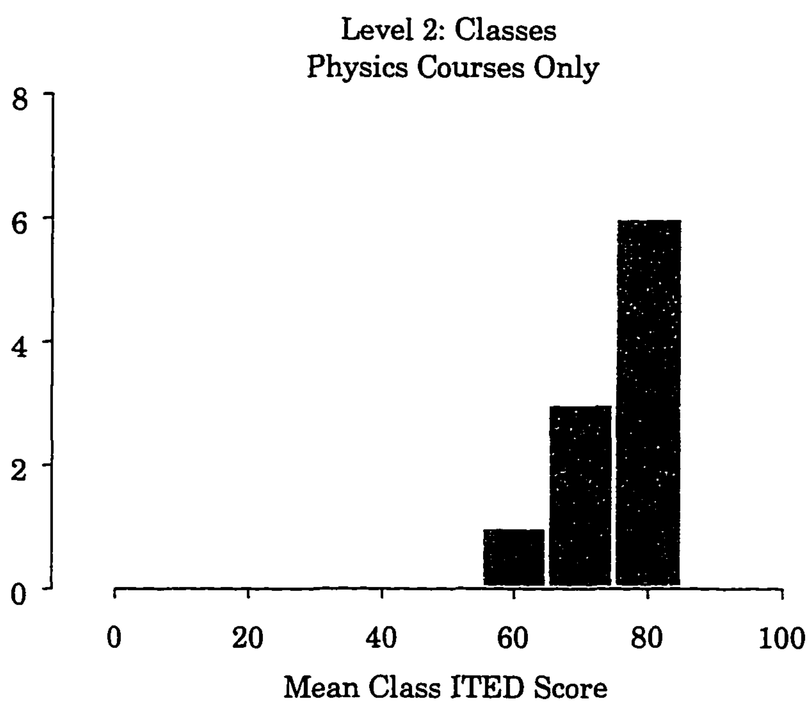
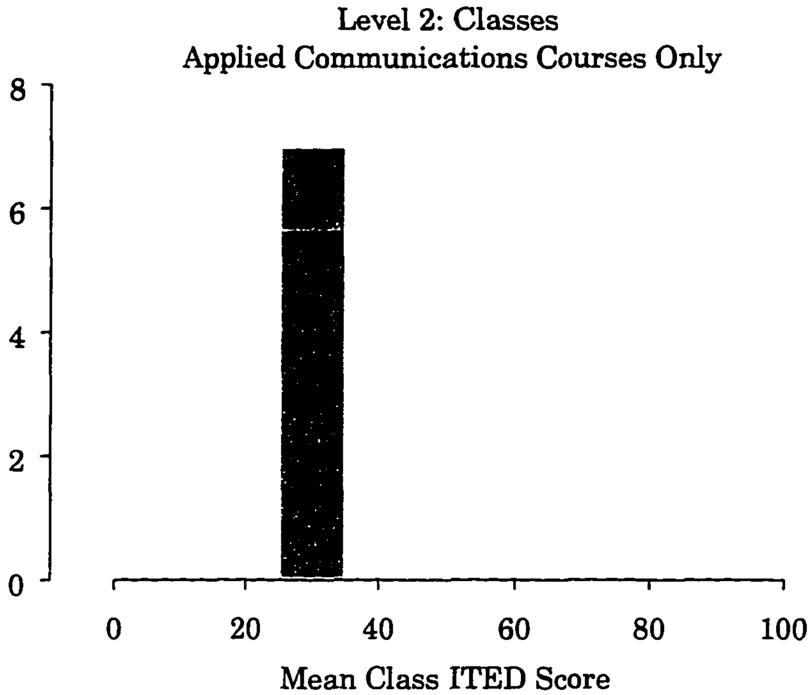


Figure C.59. Histograms comparing mean class ITED score of Principles of Technology versus Physics classes (vector data)



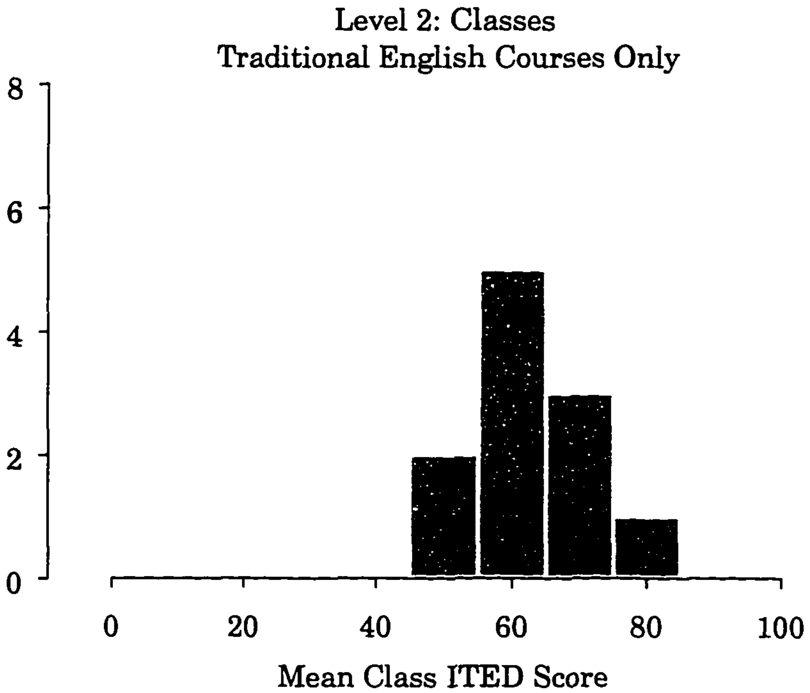
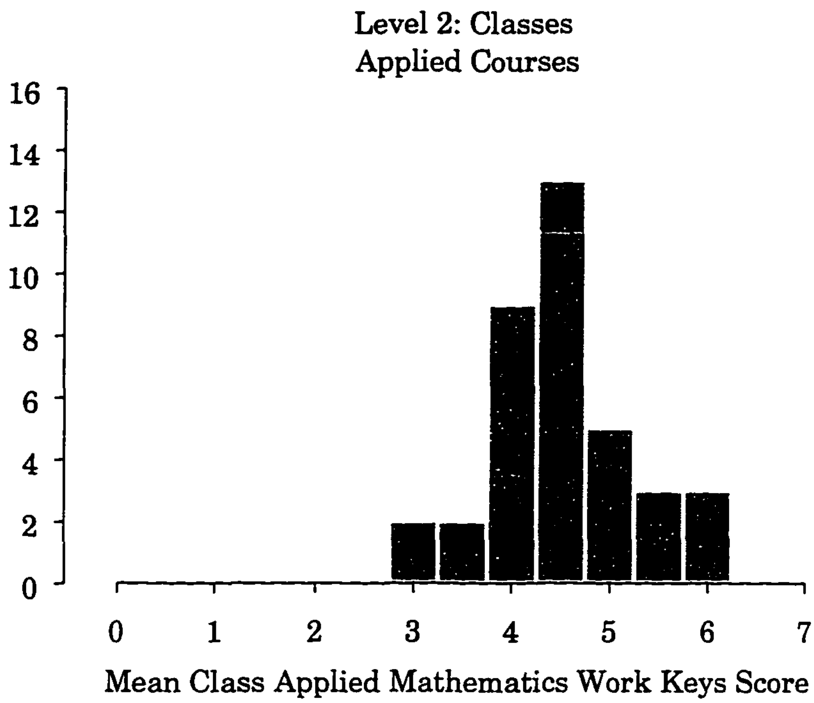


Figure C.60. Histograms comparing mean class ITED score of Applied Communications versus English classes (vector data)



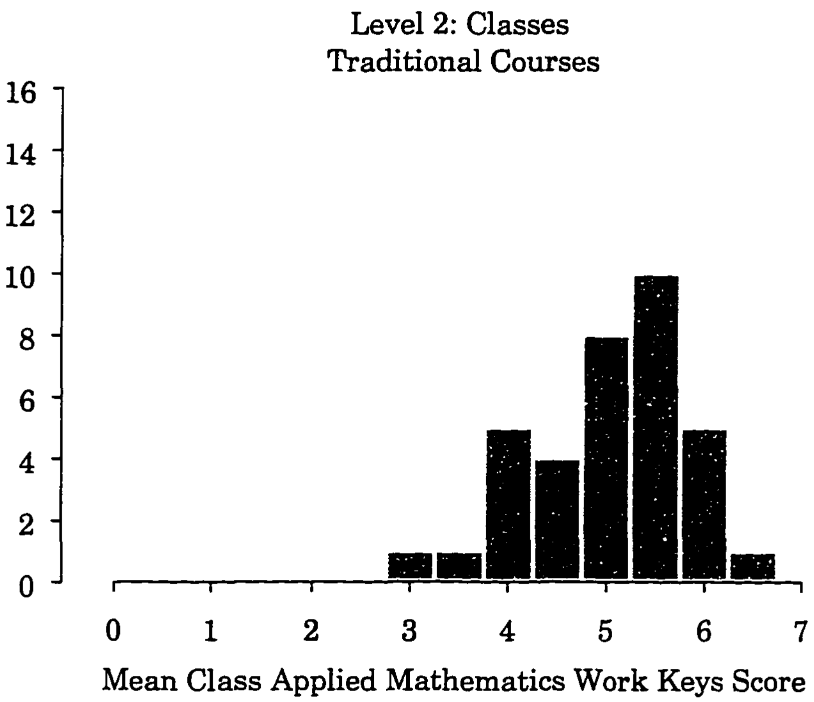


Figure C.61. Histograms comparing mean class AM Work Keys score of applied classes versus traditional classes (vector data)

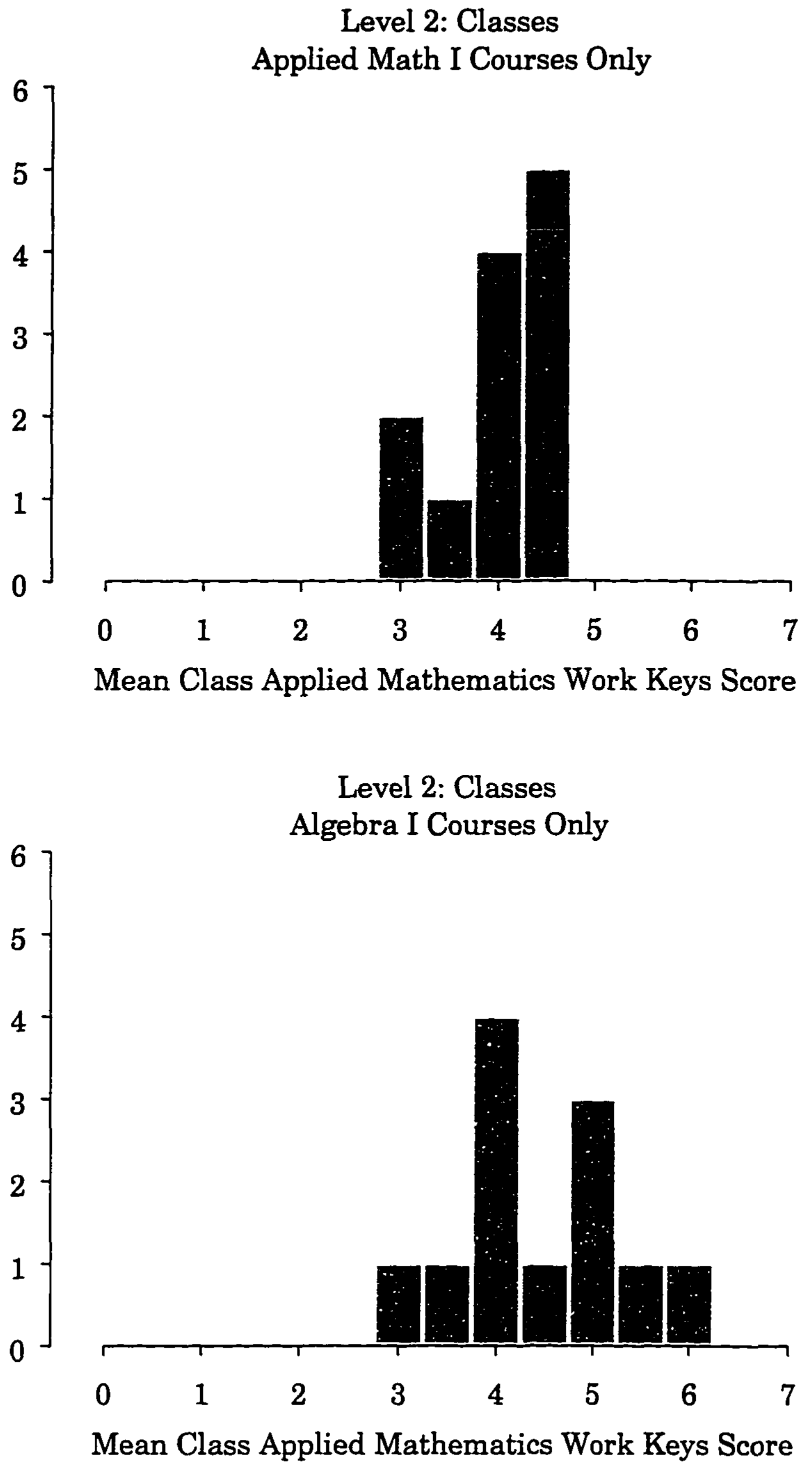


Figure C.62. Histograms comparing mean class AM Work Keys score of Applied Math I classes versus Algebra I classes (vector data)

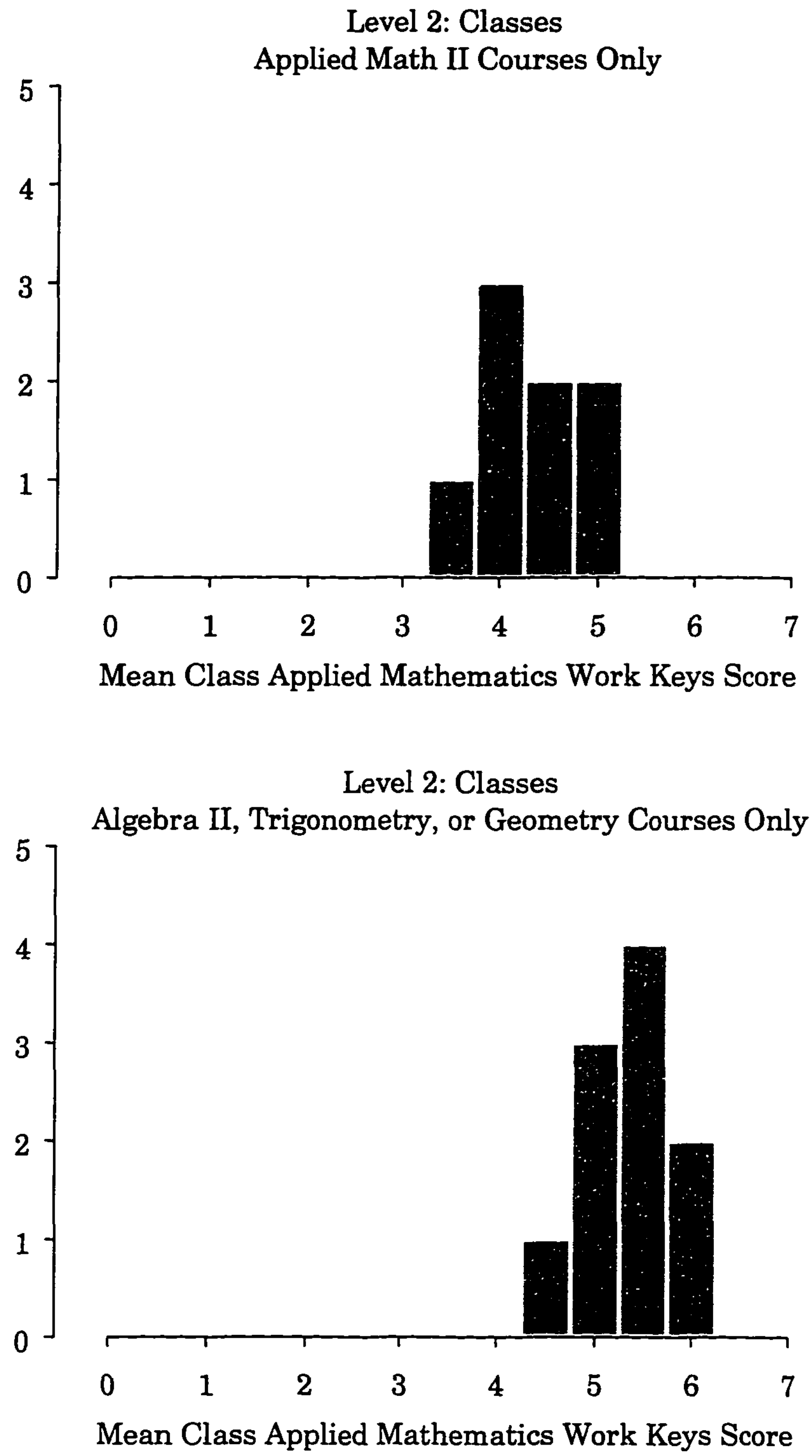
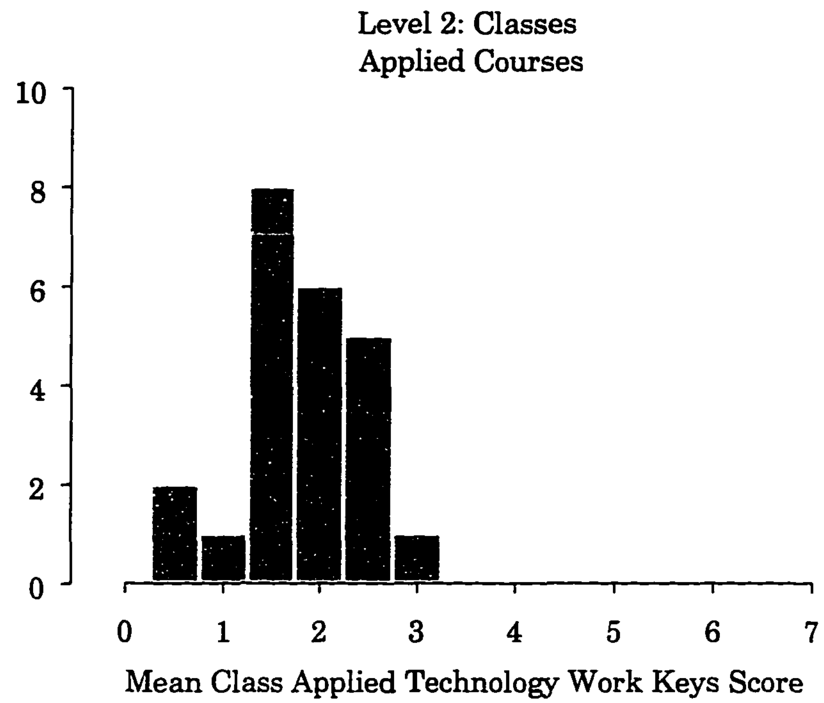


Figure C.63. Histograms comparing mean class AM Work Keys score of Applied Math II classes versus traditional Math II classes (vector data)



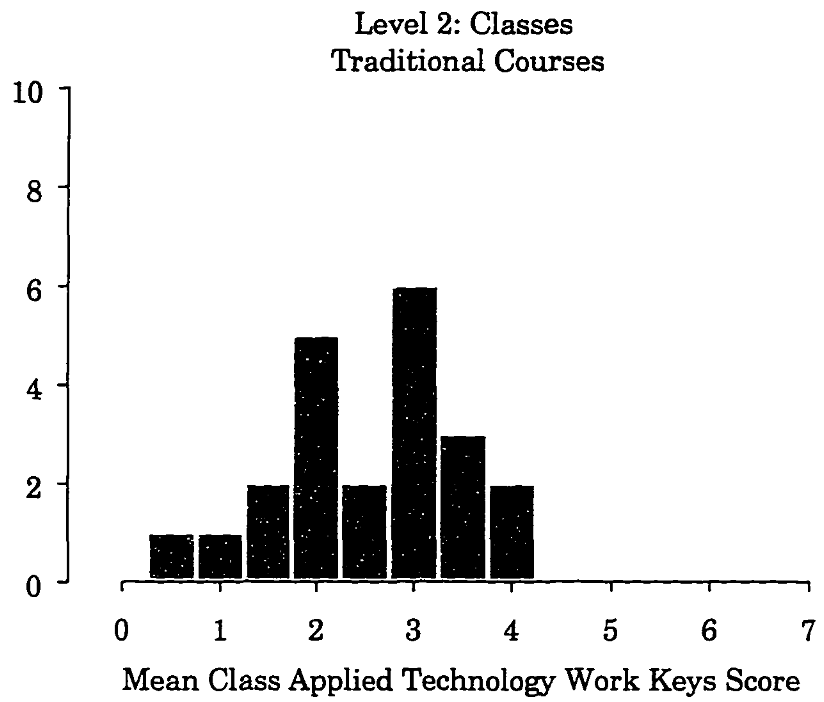
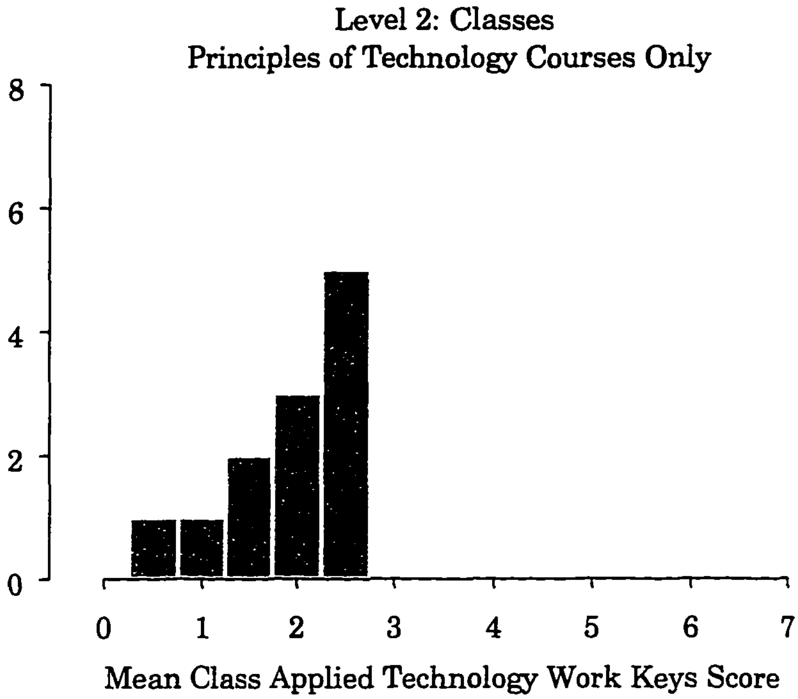


Figure C.64. Histograms comparing mean class AT Work Keys score of applied versus traditional classes (vector data)



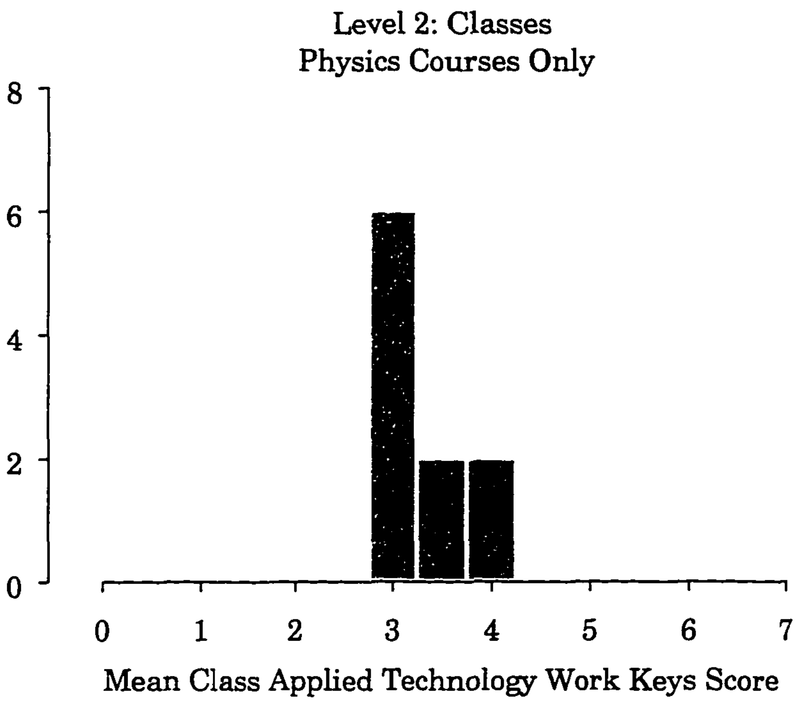
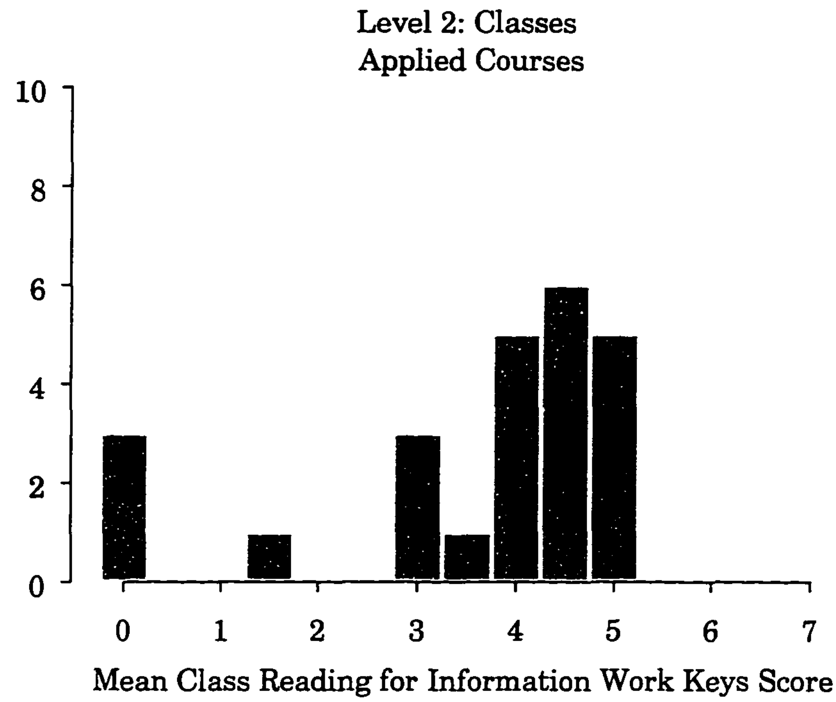


Figure C.65. Histograms comparing mean class AT Work Keys score of Principles of Technology versus Physics classes (vector data)



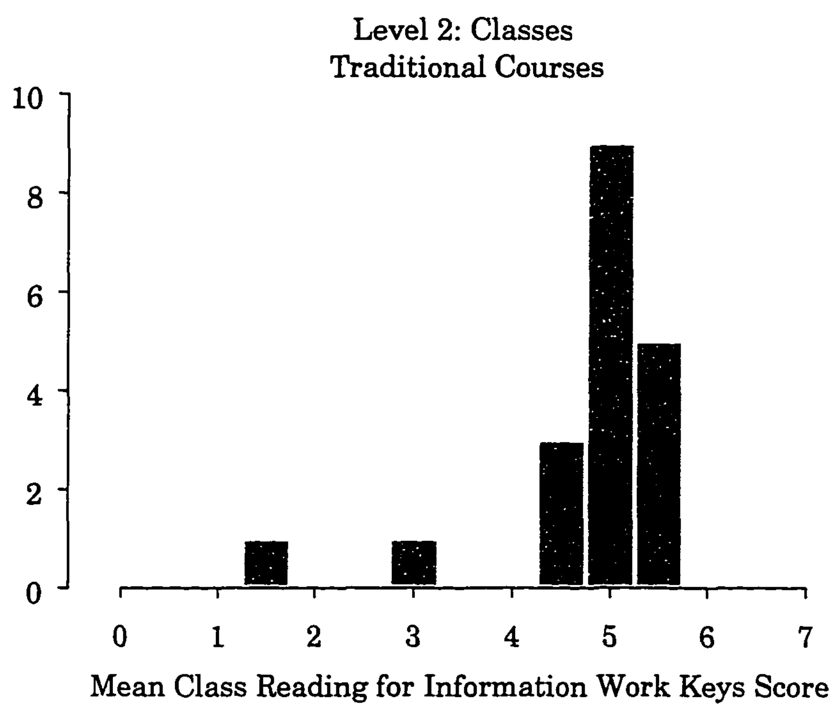
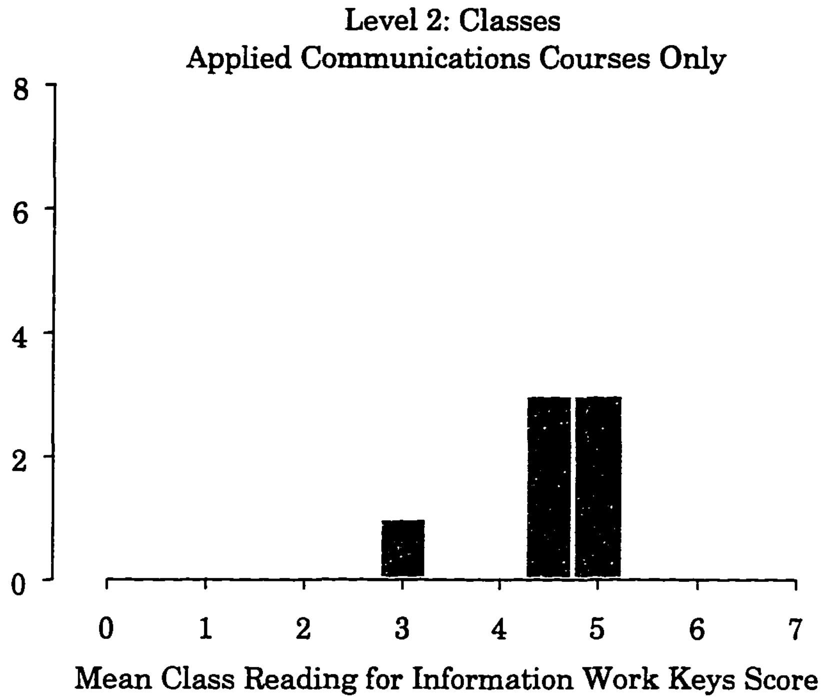


Figure C.66. Histograms comparing mean class RFI Work Keys score of applied versus traditional classes (vector data)



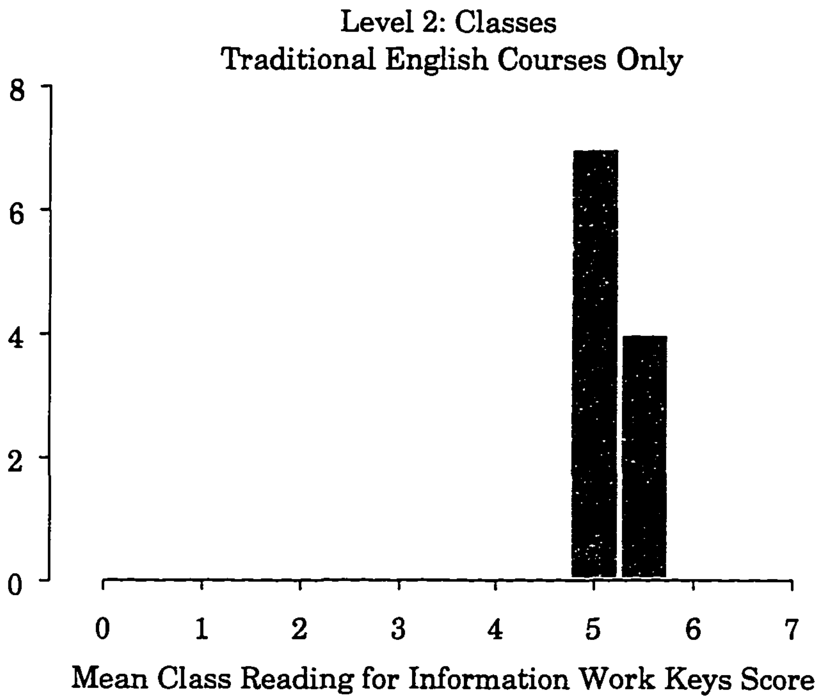


Figure C.67. Histograms comparing mean class RFI Work Keys score of Applied Communications versus English classes (vector data)

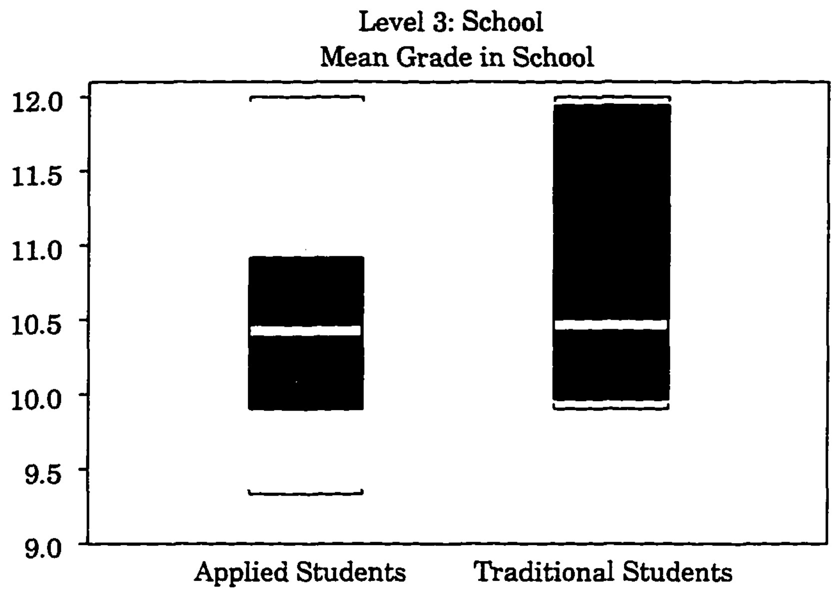


Figure C.68. Boxplot comparing Level 3 mean grade in school for applied versus traditional students (vector data)

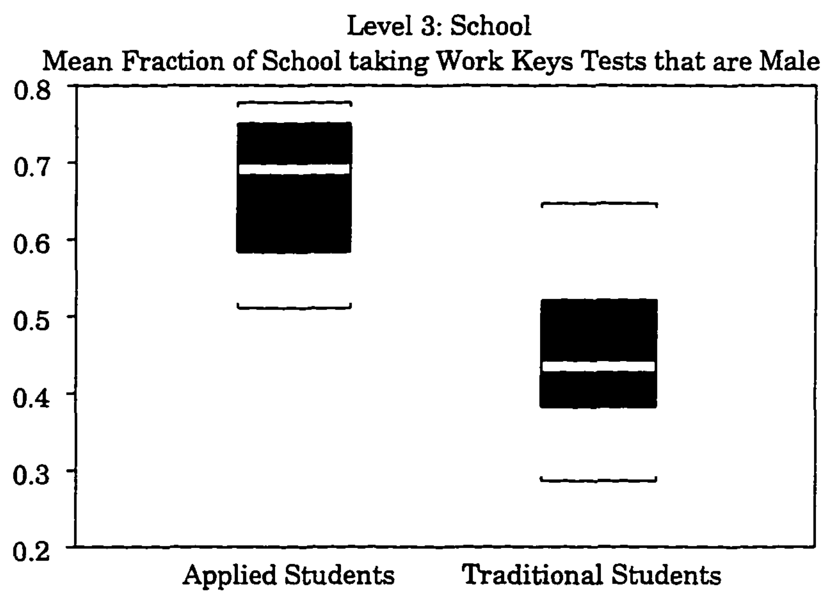


Figure C.69. Boxplot comparing Level 3 mean fraction of students taking tests who are male for applied versus traditional students (vector data)

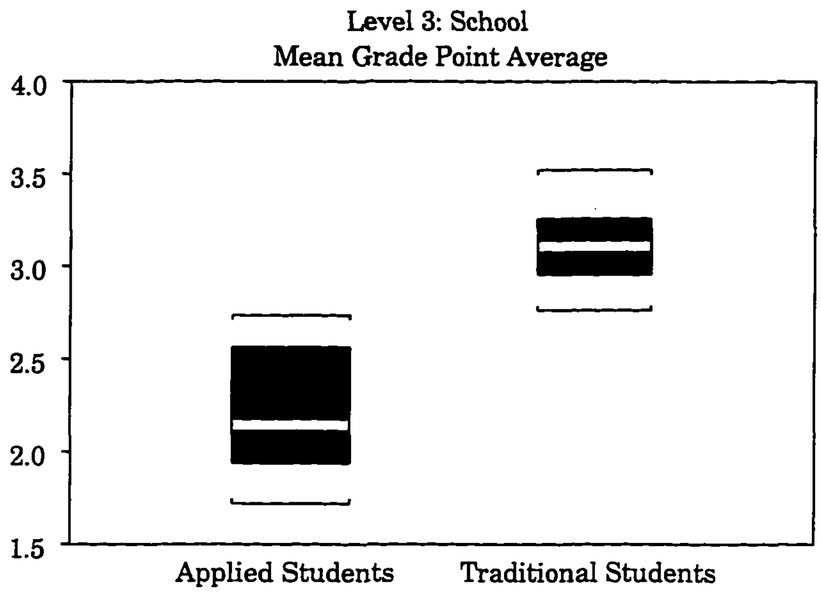


Figure C.70. Boxplot comparing Level 3 mean grade point average for applied versus traditional students (vector data)

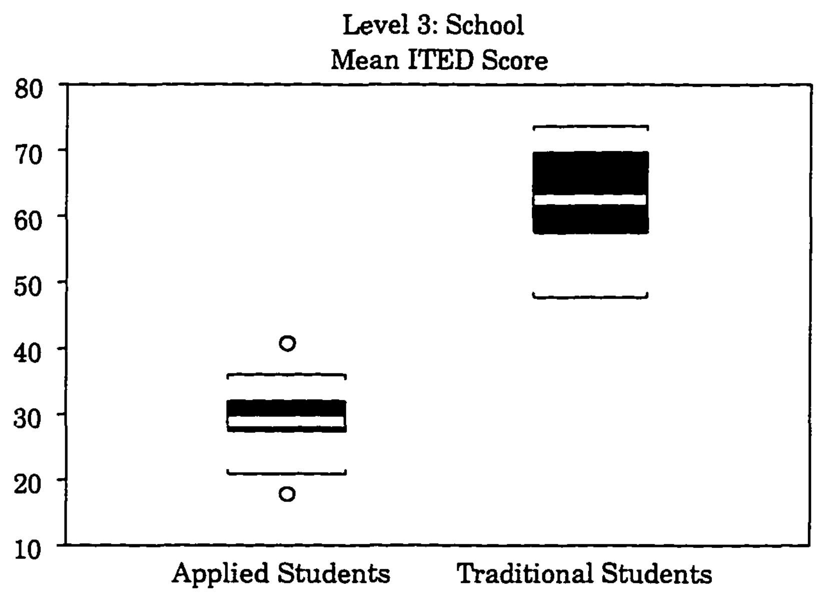


Figure C.71. Boxplot comparing Level 3 mean ITED score for applied versus traditional students (vector data)

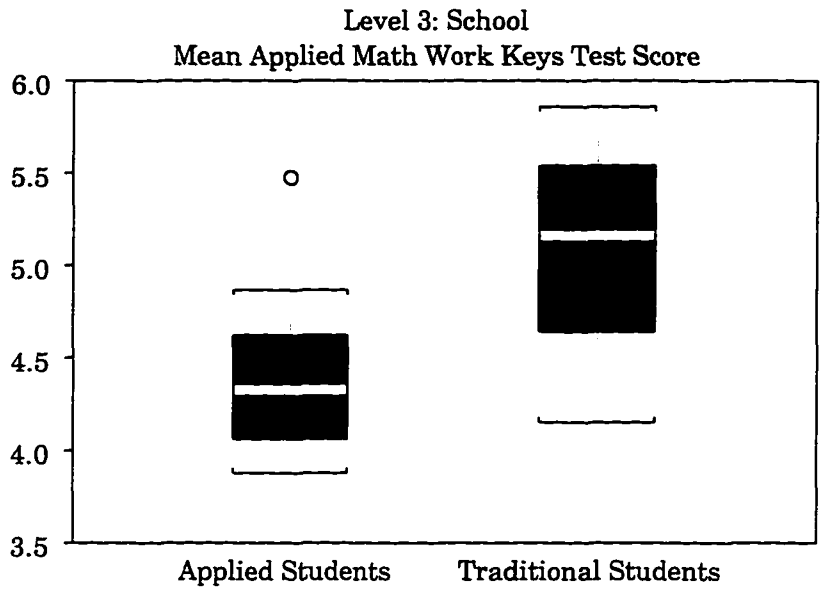


Figure C.72. Boxplot comparing Level 3 mean Applied Math Work Keys test score for applied versus traditional students (vector data)

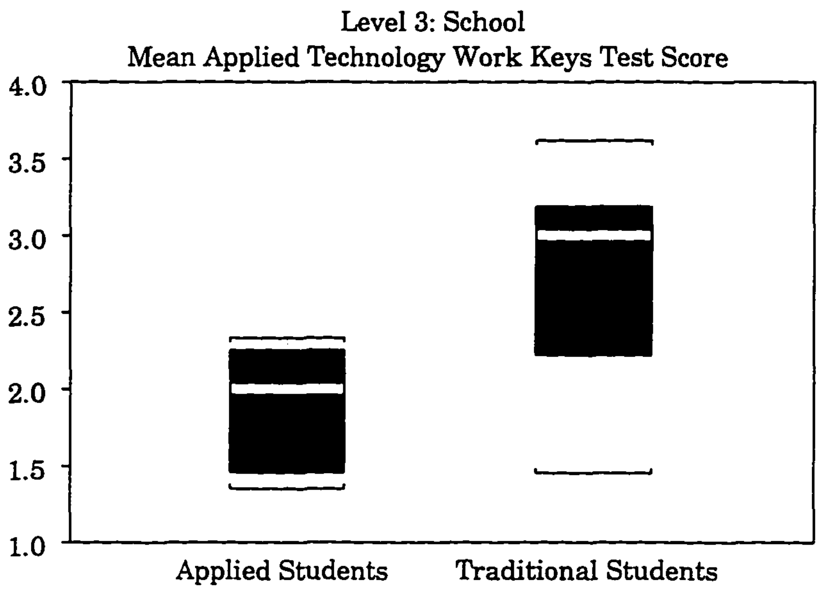


Figure C.73. Boxplot comparing Level 3 mean Applied Technology Work Keys test score for applied versus traditional students (vector data)

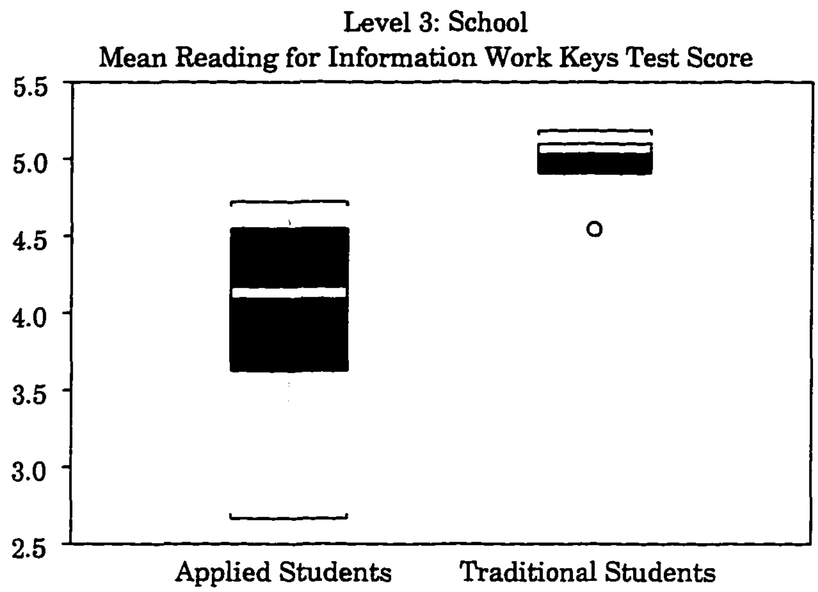


Figure C.74. Boxplot comparing Level 3 mean Reading for Information Work Keys test score for applied versus traditional students (vector data)

APPENDIX D: HLM PRINTOUTS

```
SPECIFICATIONS FOR THIS HLM RUN                          Wed Mar 12 00:27:04 1997
 --------------------------------------------------------------------------------

  Problem Title: NO TITLE

  The data source for this run = C:\HLM\DIS\CHP4MATH.SSM
  Output file name             = C:\HLM\DIS\HLM3.OUT
  The maximum number of level-2 units = 71
  The maximum number of level-3 units = 8
  The maximum number of iterations = 500
  Method of estimation: full maximum likelihood

  The outcome variable is MATHSCOR

  The model specified for the fixed effects was:
 ----------------------------------------------------

   Level-1                  Level-2          Level-3
   Coefficients             Predictors       Predictors
 ---------------------   ---------------  ----------------
        INTRCPT1, P0        INTRCPT2, B00    INTRCPT3, G000
#   GENDER slope, P1     # INTRCPT2, B10    INTRCPT3, G100
#    GRADE slope, P2     # INTRCPT2, B20    INTRCPT3, G200
#*  ACHIEV slope, P3     # INTRCPT2, B30    INTRCPT3, G300

 '#' - The residual parameter variance for the parameter has been set to zero
 '*' - This variable has been centered around its group mean

 Summary of the model specified (in equation format)
 ---------------------------------------------------

Level-1 Model
       Y = P0 + P1*(GENDER) + P2*(GRADE) + P3*(ACHIEV) + E

Level-2 Model
        P0 = B00 + R0
        P1 = B10
        P2 = B20
        P3 = B30

Level-3 Model
        B00 = G000 + U0
        B10 = G100
        B20 = G200
        B30 = G300
```

```
******* ITERATION 22 *******

 Sigma_squared =       1.00701

 Standard Error of Sigma_squared =      0.06218

Tau(pi)
 INTRCPT1,P0      0.32485


Standard Errors of Tau(pi)
 INTRCPT1,P0      0.08757


Tau(pi) (as correlations)
 INTRCPT1,P0  1.000

 ----------------------------------------------------
  Random level-1 coefficient   Reliability estimate
 ----------------------------------------------------
  INTRCPT1, P0                        0.640

Tau(beta)
 INTRCPT1
 INTRCPT2,B00
    0.08875

Tau(beta) (as correlations)
 INTRCPT1/INTRCPT2,B00  1.000

Standard Errors of Tau(beta)
 INTRCPT1
 INTRCPT2,B00
    0.07641

 ----------------------------------------------------
  Random level-2 coefficient   Reliability estimate
 ----------------------------------------------------
  INTRCPT1/INTRCPT2, B00              0.574

The value of the likelihood function at iteration 22 = -8.842150E+002
```

The outcome variable is MATHSCOR

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard Error	T-ratio	P-value
For INTRCPT1, P0				
For INTRCPT2, B00				
INTRCPT3, G000	3.234410	0.803232	4.027	0.006
For GENDER slope, P1				
For INTRCPT2, B10				
INTRCPT3, G100	0.358621	0.089220	4.020	0.000
For GRADE slope, P2				
For INTRCPT2, B20				
INTRCPT3, G200	0.131272	0.076097	1.725	0.084
For ACHIEV slope, P3				
For INTRCPT2, B30				
INTRCPT3, G300	0.030597	0.002801	10.925	0.000

Final estimation of level-1 and level-2 variance components:

Random Effect		Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1,	R0	0.56996	0.32485	63	164.47027	0.000
level-1,	E	1.00350	1.00701			

Final estimation of level-3 variance components:

Random Effect	Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1/INTRCPT2, U0	0.29790	0.08875	7	19.63876	0.007

Statistics for current covariance components model

Deviance = 1768.429985
Number of estimated parameters = 7

```
SPECIFICATIONS FOR THIS HLM RUN                      Wed Mar 12 00:39:56 1997
--------------------------------------------------------------------------------

Problem Title: NO TITLE

The data source for this run = C:\HLM\DIS\CHP4MATH.SSM
Output file name             = C:\HLM\DIS\HLM3.OUT
The maximum number of level-2 units = 71
The maximum number of level-3 units = 8
The maximum number of iterations = 500
Method of estimation: full maximum likelihood

The outcome variable is MATHSCOR

The model specified for the fixed effects was:
----------------------------------------------------

 Level-1                   Level-2            Level-3
 Coefficients              Predictors         Predictors
---------------------  ---------------  ----------------
      INTRCPT1. P0        INTRCPT2. B00    INTRCPT3. G000
#  GENDER slope. P1     # INTRCPT2. B10    INTRCPT3. G100
#* ACHIEV slope. P2     # INTRCPT2. B20    INTRCPT3. G200

'#' - The residual parameter variance for the parameter has been set to zero
'*' - This variable has been centered around its group mean

Summary of the model specified (in equation format)
----------------------------------------------------

Level-1 Model
       Y = P0 + P1*(GENDER) + P2*(ACHIEV) + E

Level-2 Model
        P0 = B00 + R0
        P1 = B10
        P2 = B20

Level-3 Model
        B00 = G000 + U0
        B10 = G100
        B20 = G200
```

```
******* ITERATION 7 *******

 Sigma_squared =      1.00215

 Standard Error of Sigma_squared =      0.06193

Tau(pi)
 INTRCPT1,P0      0.33411


Standard Errors of Tau(pi)
 INTRCPT1,P0      0.08953


Tau(pi) (as correlations)
 INTRCPT1,P0  1.000

 ----------------------------------------------------
  Random level-1 coefficient   Reliability estimate
 ----------------------------------------------------
  INTRCPT1, P0                       0.646

Tau(beta)
 INTRCPT1
 INTRCPT2,B00
    0.17434

Tau(beta) (as correlations)
 INTRCPT1/INTRCPT2,B00  1.000

Standard Errors of Tau(beta)
 INTRCPT1
 INTRCPT2,B00
    0.12158

 ----------------------------------------------------
  Random level-2 coefficient   Reliability estimate
 ----------------------------------------------------
  INTRCPT1/INTRCPT2, B00               0.712

The value of the likelihood function at iteration 7 = -8.852294E+002
```

The outcome variable is MATHSCOR

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard Error	T-ratio	P-value
For INTRCPT1, P0				
For INTRCPT2, B00				
INTRCPT3, G000	4.599604	0.181635	25.323	0.000
For GENDER slope, P1				
For INTRCPT2, B10				
INTRCPT3, G100	0.362078	0.089104	4.064	0.000
For ACHIEV slope, P2				
For INTRCPT2, B20				
INTRCPT3, G200	0.029522	0.002724	10.839	0.000

Final estimation of level-1 and level-2 variance components:

Random Effect		Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1,	R0	0.57802	0.33411	63	138.80340	0.000
level-1,	E	1.00107	1.00215			

Final estimation of level-3 variance components:

Random Effect	Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1/INTRCPT2, U0	0.41753	0.17434	7	30.55764	0.000

Statistics for current covariance components model

Deviance = 1770.458889
Number of estimated parameters = 6

SVariance-covariance components test

Chi-squared statistic = 2.028958
Number of degrees of freedom = 1
P-value = 0.150435

```
  SPECIFICATIONS FOR THIS HLM RUN                        Wed Mar 12 00:54:39 1997
-------------------------------------------------------------------------------

  Problem Title: NO TITLE

  The data source for this run = C:\HLM\DIS\CHP4MATH.SSM
  Output file name             = C:\HLM\DIS\HLM3.OUT
  The maximum number of level-2 units = 71
  The maximum number of level-3 units = 8
  The maximum number of iterations = 500
  Method of estimation: full maximum likelihood

The outcome variable is MATHSCOR

  The model specified for the fixed effects was:
 ----------------------------------------------------

   Level-1                Level-2          Level-3
   Coefficients           Predictors       Predictors
 ---------------------  ---------------  ----------------
        INTRCPT1, P0      INTRCPT2, B00   INTRCPT3, G000
#   GENDER slope, P1    # INTRCPT2, B10   INTRCPT3, G100
#*  ACHIEV slope, P2      INTRCPT2, B20   INTRCPT3, G200

 '#' - The residual parameter variance for the parameter has been set to zero
 '*' - This variable has been centered around its group mean

 Summary of the model specified (in equation format)
 ---------------------------------------------------
Level-1 Model
       Y = P0 + P1*(GENDER) + P2*(ACHIEV) + E

Level-2 Model
        P0 = B00 + R0
        P1 = B10
        P2 = B20

Level-3 Model
        B00 = G000 + U0
        B10 = G100
        B20 = G200 + U1
```

```
******* ITERATION 17 *******

 Sigma_squared =        0.98289

 Standard Error of Sigma_squared =      0.06112

Tau(pi)
 INTRCPT1.P0       0.33219


Standard Errors of Tau(pi)
 INTRCPT1.P0       0.08854


Tau(pi) (as correlations)
 INTRCPT1.P0  1.000

 ----------------------------------------------------
  Random level-1 coefficient   Reliability estimate
 ----------------------------------------------------
  INTRCPT1, P0                          0.649

Tau(beta)
 INTRCPT1       ACHIEV
 INTRCPT2.B00 INTRCPT2.B20
    0.18545      0.00351
    0.00351      0.00010

Tau(beta) (as correlations)
 INTRCPT1/INTRCPT2.B00  1.000  0.829
   ACHIEV/INTRCPT2.B20  0.829  1.000

Standard Errors of Tau(beta)
 INTRCPT1       ACHIEV
 INTRCPT2.B00 INTRCPT2.B20
    0.12605      0.00259
    0.00259      0.00008

 ----------------------------------------------------
  Random level-2 coefficient   Reliability estimate
 ----------------------------------------------------
  INTRCPT1/INTRCPT2, B00                0.722
    ACHIEV/INTRCPT2, B20                0.573

The value of the likelihood function at iteration 17 = -8.822622E+002
```

The outcome variable is MATHSCOR

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard Error	T-ratio	P-value
For INTRCPT1, P0				
For INTRCPT2, B00				
INTRCPT3, G000	4.595450	0.184533	24.903	0.000
For GENDER slope, P1				
For INTRCPT2, B10				
INTRCPT3, G100	0.362815	0.088733	4.089	0.000
For ACHIEV slope, P2				
For INTRCPT2, B20				
INTRCPT3, G200	0.030483	0.004518	6.746	0.000

Final estimation of level-1 and level-2 variance components:

Random Effect		Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1,	R0	0.57636	0.33219	63	164.11313	0.000
level-1,	E	0.99141	0.98289			

Final estimation of level-3 variance components:

Random Effect	Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1/INTRCPT2, U0	0.43064	0.18545	7	30.83316	0.000
ACHIEV/INTRCPT2, U1	0.00983	0.00010	7	17.31933	0.015

Statistics for current covariance components model

Deviance = 1764.524422
Number of estimated parameters = 8

Variance-covariance components test

Chi-squared statistic = 5.934441
Number of degrees of freedom = 2
P-value = 0.049892

Exploratory Analysis: estimated level-2 coefficients and their standard errors obtained by regressing EB residuals on level-2 predictors selected for possible inclusion in subsequent HLM runs

Level-1 Coefficient	Potential Level-2 Predictors				
	TYPE	RELVNT	CMACHIEV	CMGENDER	CMGRADE
INTRCPT1. P0					
Coefficient	0.452	-0.309	0.019	0.155	0.082
Standard Error	0.094	0.175	0.003	0.221	0.097
t value	4.825	-1.764	7.108	0.701	0.844

```
SPECIFICATIONS FOR THIS HLM RUN                      Wed Mar 12 01:01:45 1997
-------------------------------------------------------------------------------

 Problem Title: NO TITLE

 The data source for this run = C:\HLM\DIS\CHP4MATH.SSM
 Output file name             = C:\HLM\DIS\HLM3.OUT
 The maximum number of level-2 units = 71
 The maximum number of level-3 units = 8
 The maximum number of iterations = 500
 Method of estimation: full maximum likelihood

 The outcome variable is MATHSCOR

 The model specified for the fixed effects was:
 ----------------------------------------------------
   Level-1                  Level-2           Level-3
   Coefficients             Predictors        Predictors
 --------------------   ---------------   ----------------
        INTRCPT1, P0       INTRCPT2, B00     INTRCPT3, G000
#  GENDER slope, P1     # INTRCPT2, B10     INTRCPT3, G100
#   GRADE slope, P2     # INTRCPT2, B20     INTRCPT3, G200
#* ACHIEV slope, P3       INTRCPT2, B30     INTRCPT3, G300

 '#' - The residual parameter variance for the parameter has been set to zero
 '*' - This variable has been centered around its group mean

 Summary of the model specified (in equation format)
 ---------------------------------------------------
Level-1 Model
       Y = P0 + P1*(GENDER) + P2*(GRADE) + P3*(ACHIEV) + E

Level-2 Model
        P0 = B00 + R0
        P1 = B10
        P2 = B20
        P3 = B30

Level-3 Model
        B00 = G000 + U0
        B10 = G100
        B20 = G200
        B30 = G300 + U1
```

```
******* ITERATION 22 *******

 Sigma_squared =      0.98658

 Standard Error of Sigma_squared =      0.06134

Tau(pi)
 INTRCPT1,P0      0.32635


Standard Errors of Tau(pi)
 INTRCPT1,P0      0.08735


Tau(pi) (as correlations)
 INTRCPT1,P0  1.000

 ----------------------------------------------------
  Random level-1 coefficient   Reliability estimate
 ----------------------------------------------------
  INTRCPT1, P0                        0.645

Tau(beta)
 INTRCPT1       ACHIEV
 INTRCPT2,B00 INTRCPT2,B30
    0.12938       0.00275
    0.00275       0.00009

Tau(beta) (as correlations)
 INTRCPT1/INTRCPT2,B00  1.000  0.819
   ACHIEV/INTRCPT2,B30  0.819  1.000

Standard Errors of Tau(beta)
 INTRCPT1       ACHIEV
 INTRCPT2,B00 INTRCPT2,B30
    0.09680       0.00216
    0.00216       0.00007

 ----------------------------------------------------
  Random level-2 coefficient   Reliability estimate
 ----------------------------------------------------
  INTRCPT1/INTRCPT2, B00              0.653
    ACHIEV/INTRCPT2, B30              0.551

The value of the likelihood function at iteration 22 = -8.819268E+002
```

The outcome variable is MATHSCOR

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard Error	T-ratio	P-value
For INTRCPT1, P0				
For INTRCPT2, B00				
INTRCPT3, G000	3.808908	0.816888	4.663	0.002
For GENDER slope, P1				
For INTRCPT2, B10				
INTRCPT3, G100	0.362097	0.088839	4.076	0.000
For GRADE slope, P2				
For INTRCPT2, B20				
INTRCPT3, G200	0.075829	0.077258	0.981	0.327
For ACHIEV slope, P3				
For INTRCPT2, B30				
INTRCPT3, G300	0.031042	0.004429	7.008	0.000

Final estimation of level-1 and level-2 variance components:

Random Effect		Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1,	R0	0.57127	0.32635	63	175.36680	0.000
level-1,	E	0.99327	0.98658			

Final estimation of level-3 variance components:

Random Effect	Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1/INTRCPT2, U0	0.35969	0.12938	7	24.00045	0.001
ACHIEV/INTRCPT2, U1	0.00935	0.00009	7	16.40353	0.022

Statistics for current covariance components model

Deviance = 1763.853523
Number of estimated parameters = 9

Variance-covariance components test

Chi-squared statistic = 0.670891
Number of degrees of freedom = 1
P-value = >.500

APPENDIX E: CHECK ON LEVEL-1 ASSUMPTIONS

Overview

The first of the "Methodological Assumptions" listed in Chapter 3 states that, "Conditional on a student's Level-1 predictor variables, the within-class errors are normal and independent with class means of zero and equal variances across classes." The HLM software used in this investigation allows one to test the homogeneity of Level-1 variances in a two-level model. This test is based on a statistic whose formula uses a standardized measure of dispersion for each group. This statistic has a large sample χ^2 distribution under the homogeneity hypothesis, but is only appropriate when, "the data are normal and sample sizes per unit are 10 or more" (Bryk and Raudenbush, 1992, p. 208). This statistic is shown in Table E.1 for each of the three Work Keys tests, however the sample size condition was not met for the complete data set (see Table E.2). These tables and the plots that follow were produced to explore the information available from the residuals files, but care should be taken as to the inferences drawn from them because of the number of within-class sample sizes less than 10.

Figures E.1 through E.3 show EDA plots of the standardized measure of dispersion for each of the three Work Keys test groups (that is, the natural logarithms of the within-class residual standard deviations from the final fitted effects model). These plots are used to check the assumption of normality of the within-class errors.

Figures E.4 through E.6 show plots of the Mahalanobis distance versus the expected values of the order statistics for a sample of size j selected from a population that is distributed $\chi^2_{(\upsilon)}$. Here one is looking at the Level-2 normality assumption. The Mahalanobis distance is the standardized squared distance of a unit from the center of a υ-dimensional distribution, where υ is the number of random effects per unit. Essentially, the Mahalanobis distance provides a single, summary measure of the distance of a unit's Empirical Bayes (EB) estimates from its "fitted value". If the normality assumption is true, then the Mahalanobis distances should be distributed approximately and plots in Figures E.4 through E.6 will resemble 45 degree lines. Note that these plots are subject to the same limitations previously mentioned regarding sample sizes; the Level-1 sample sizes should be 10 or above before one can expect these to be reasonable diagnostic tools (Bryk, Raudenbush, and Congdon, 1996, pp. 34-35).

Figures E.7 through E.9 are plots of the EB residuals versus the fitted intercept values. The fitted intercept values are shown in Tables E.3. Bryk et al. (1996) have the following to say regarding EB estimates of randomly varying Level-1 coefficients:

> These estimates of the level-1 coefficients for each unit j are optimal composites of an estimate based on the data from that unit and an estimate based on data from other similar units. Intuitively, we are borrowing strength from all of the information present in the ensemble of data to improve the level-1 coefficient estimates for each of the J units. These "EB" estimates are also referred to as "shrunken estimates" of the level-1 coefficients. (p. 4)

A more general discussion of these EB estimates is provided in the book by Bryk and Raudenbush (1992, pp. 39-44; 76-82).

One other note: The statistics for the three-level Applied Mathematics and Reading for Information models were obtained by combining class and school levels before testing for within-class variance homogeneity. Combining Levels 2 and 3 had little effect on the characteristics of Level-1 data; the results from both the two-level and the three-level models were compared prior to checking within-class variance homogeneity using the two-level model. The Level-1 homogeneity test is not available in the software for the three-level model.

Table E.1. Test for Level-1 homogeneity of variance

Work Keys Test	χ^2 Statistic	df	p-value
Applied Mathematics	117.50	62	0.000
Applied Technology	41.31	37	0.288
Reading for Information	70.02	32	0.000

Table E.2. Number of classes with sample size greater than or equal to 10

Work Keys Test	$n_j \geq 10$	% of total
Applied Mathematics	28	39.4%
Applied Technology	20	52.6%
Reading for Information	14	32.6%

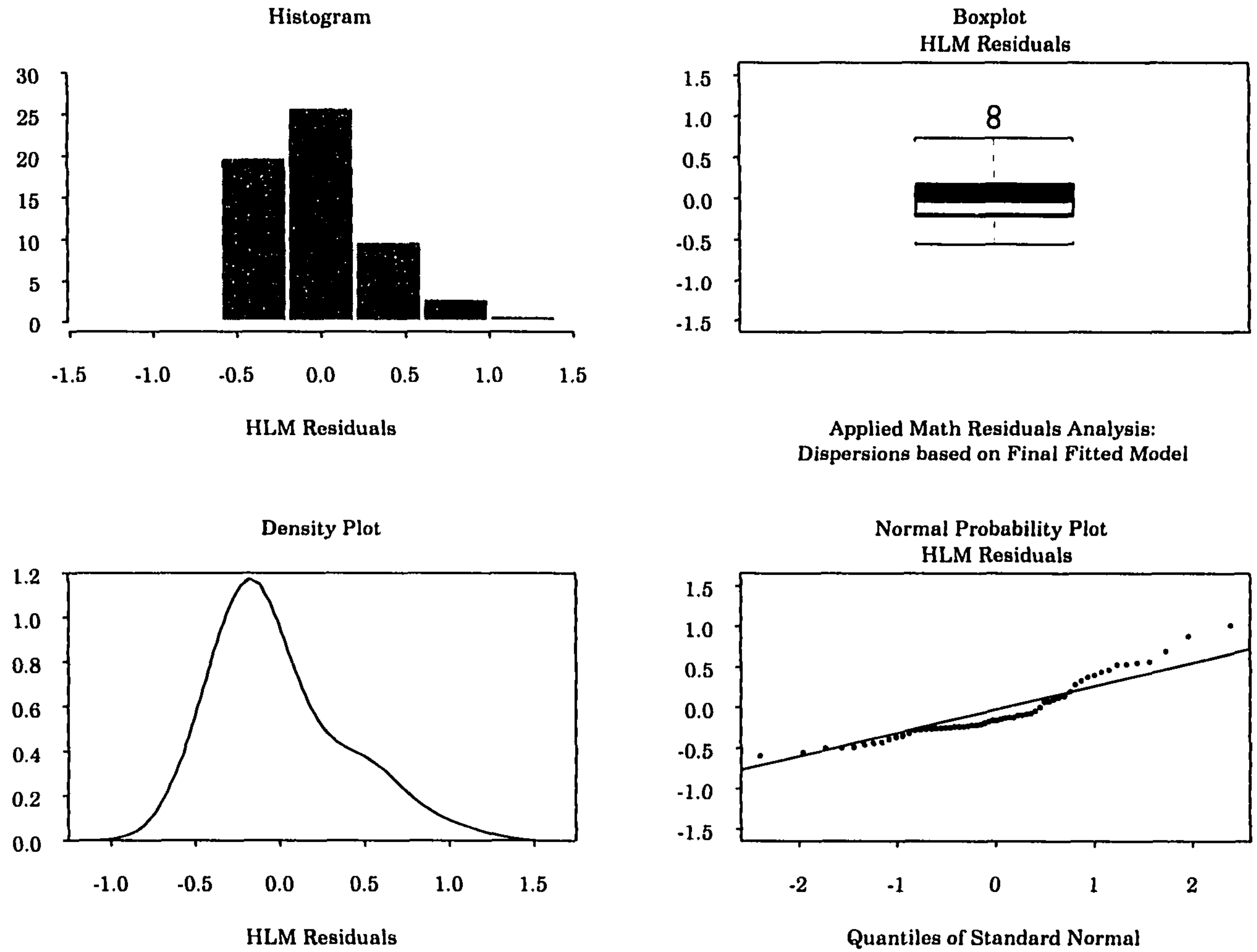


Figure E.1. Exploratory Data Analysis plots of Applied Mathematics HLM Residuals. The plotted data are the natural logarithms of the within-class residual standard deviations from the final fitted effects two-level model.

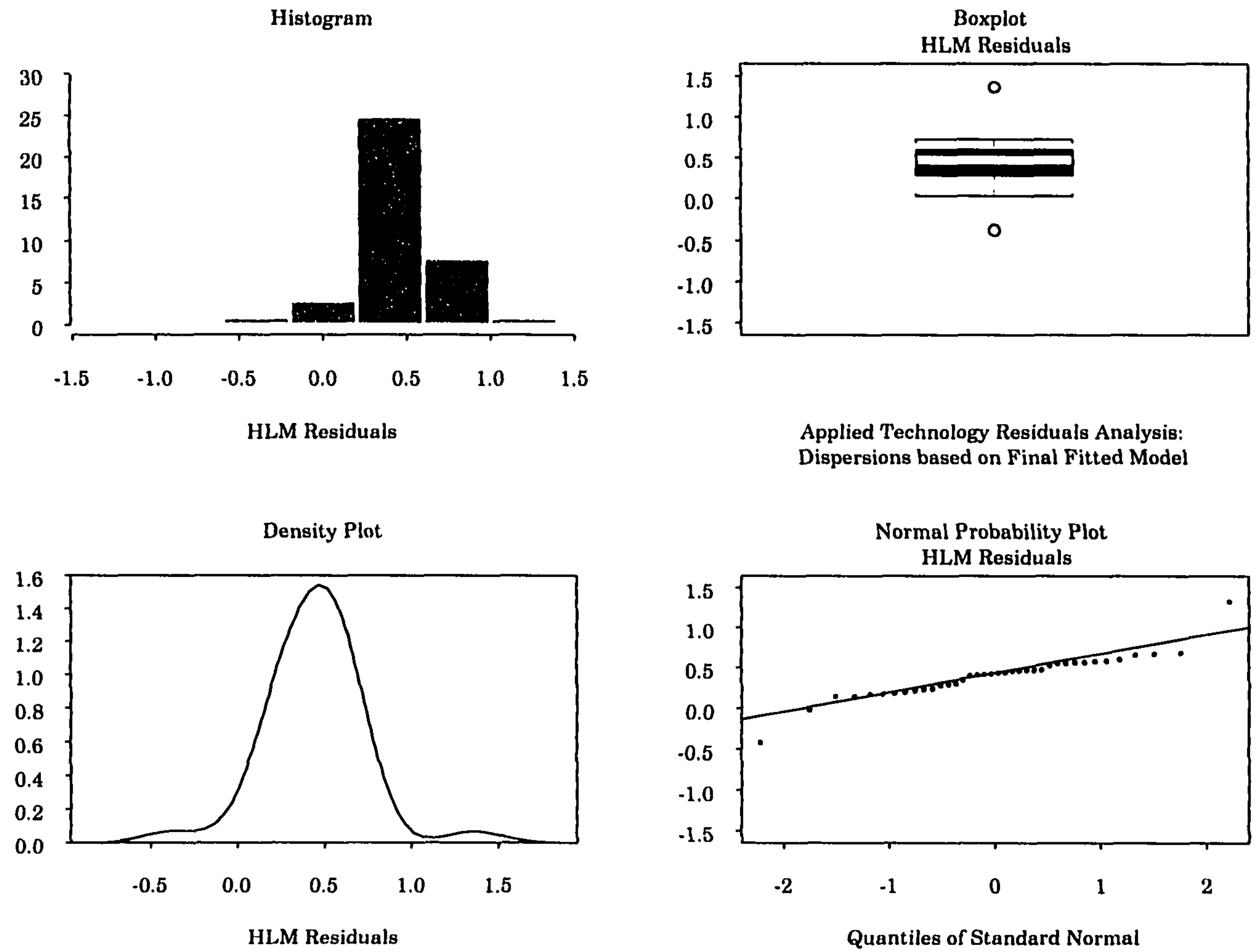


Figure E.2. Exploratory Data Analysis plots of Applied Technology HLM Residuals. The plotted data are the natural logarithms of the within-class residual standard deviations from the final fitted effects two-level model.

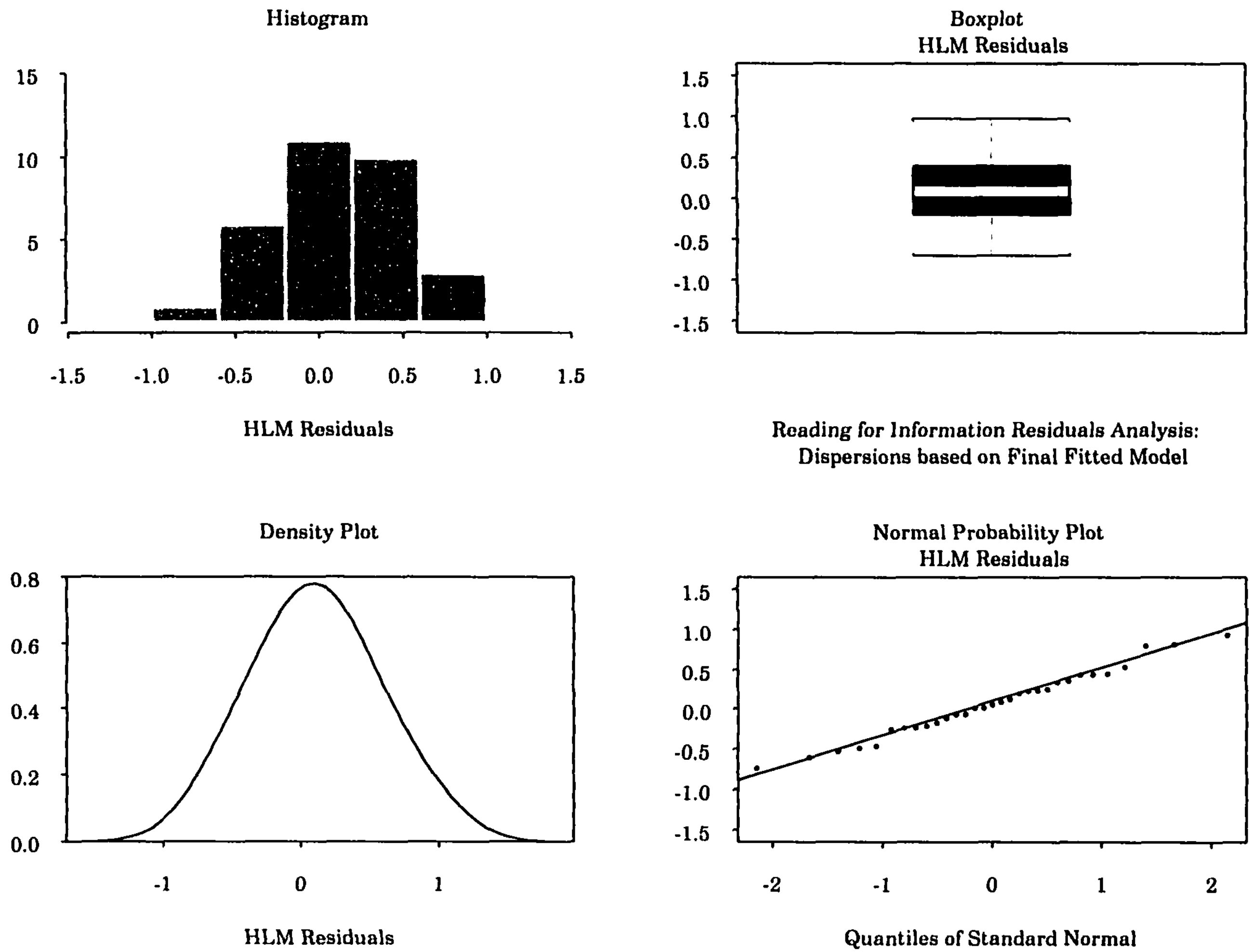


Figure E.3. Exploratory Data Analysis plots of Reading for Information HLM Residuals. The plotted data are the natural logarithms of the within-class residual standard deviations from the final fitted effects two-level model.

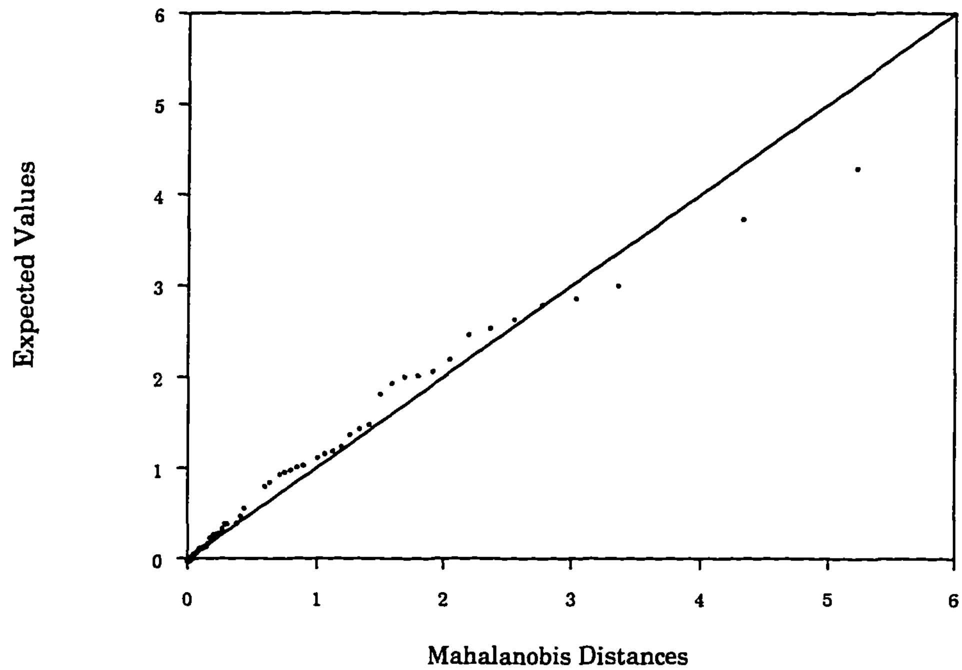


Figure E.4. Mahalanobis plot for examining Applied Mathematics Level-2 normality assumptions

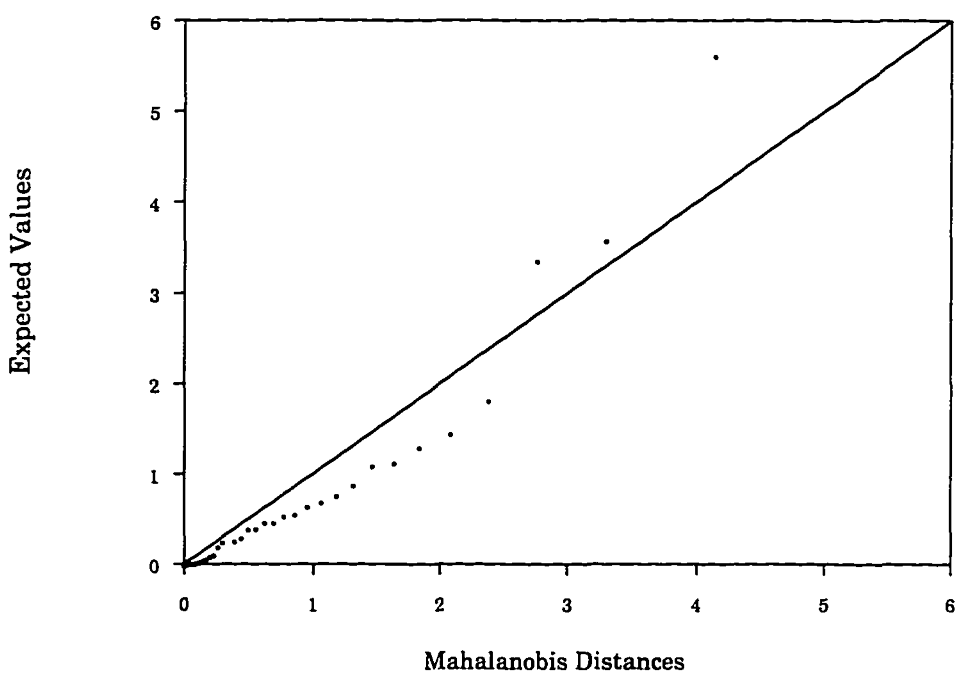


Figure E.5. Mahalanobis plot for examining Applied Technology Level-2 normality assumptions

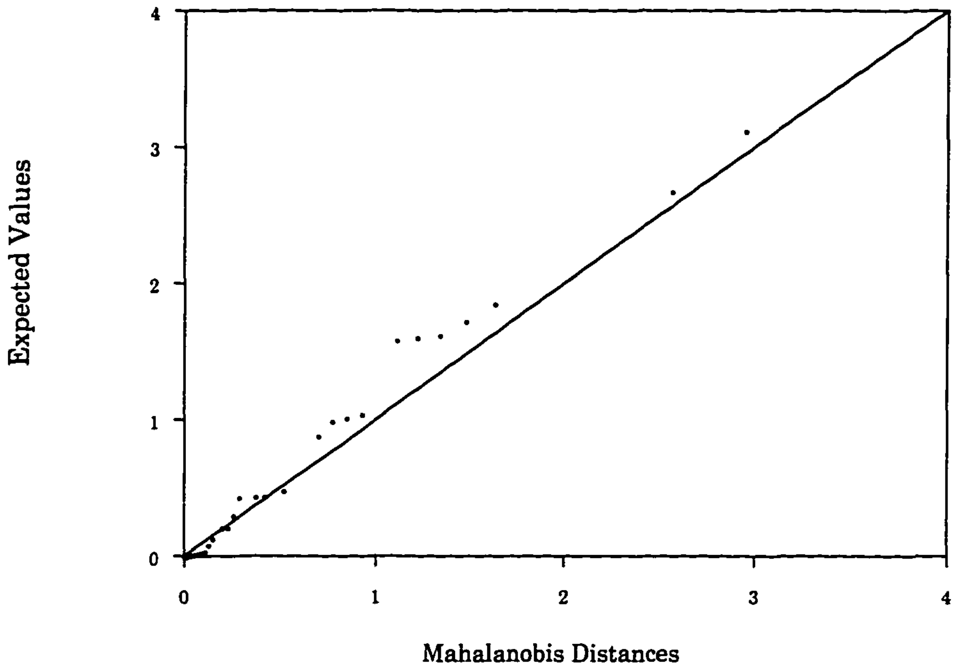


Figure E.6. Mahalanobis plot for examining Reading for Information Level-2 normality assumptions

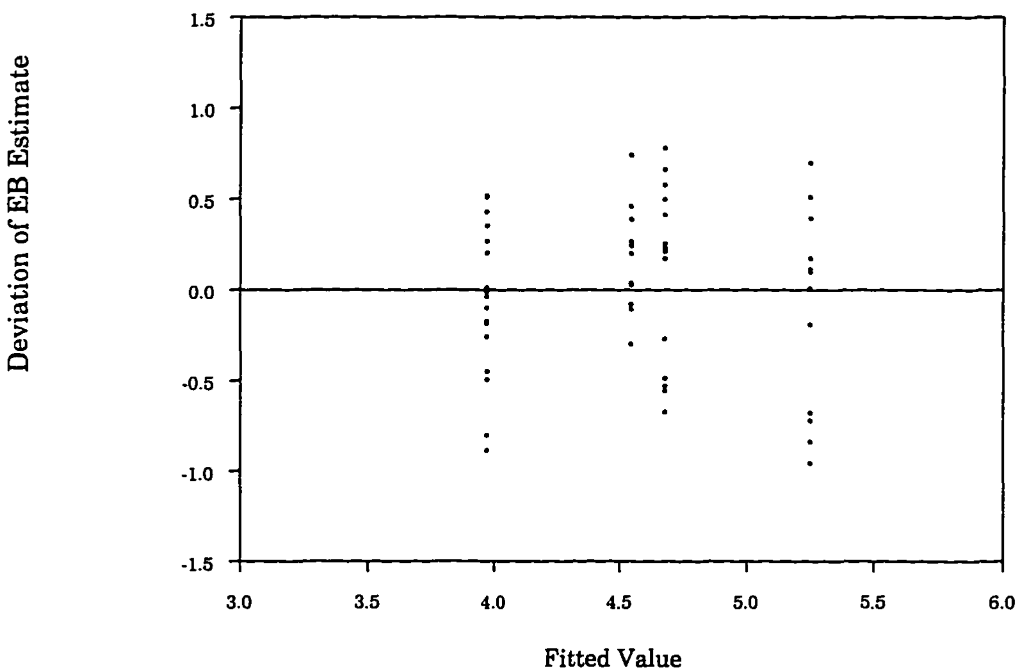


Figure E.7. Deviation of the Applied Mathematics EB estimate of the randomly varying Level-1 intercept from its predicted value based on the Level-2 model

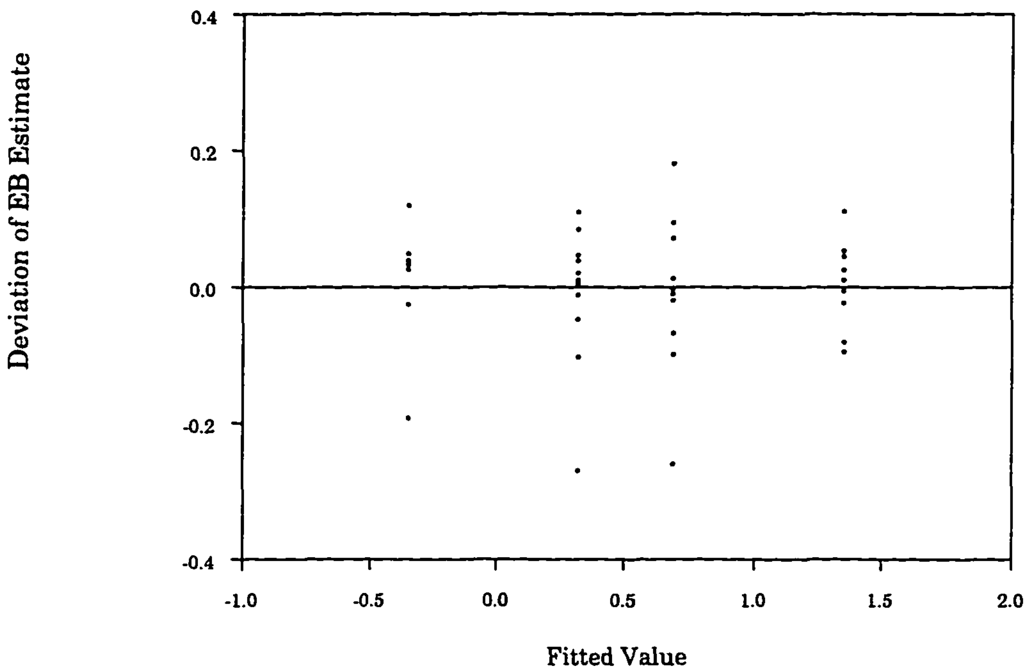


Figure E.8. Deviation of the Applied Technology EB estimate of the randomly varying Level-1 intercept from its predicted value based on the Level-2 model

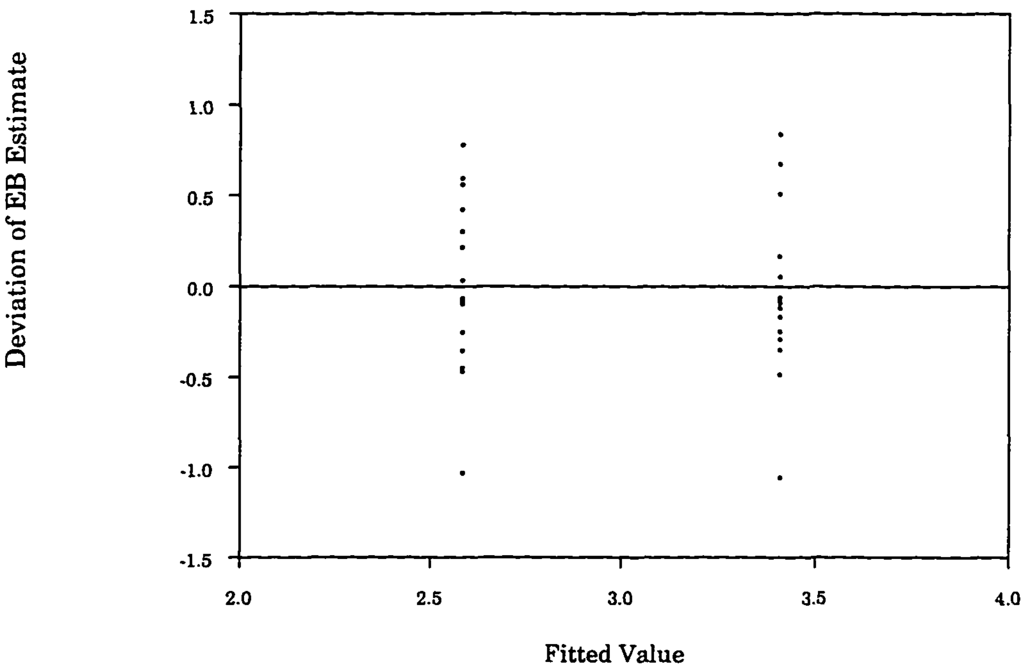


Figure E.9. Deviation of the Reading for Information EB estimate of the randomly varying Level-1 intercept from its predicted value based on the Level-2 model

Table E.3. Fitted intercept value

Work Keys Test	Fitted Intercept	Curricula Type	Relevancy Variable
Applied Mathematics	3.971	applied	relevant
	4.545	applied	not relevant
	4.677	traditional	relevant
	5.251	traditional	not relevant
Applied Technology	0.320	applied	relevant
	-0.345	applied	not relevant
	1.355	traditional	relevant
	0.690	traditional	not relevant
Reading for Information	2.585	applied	---
	3.408	traditional	---

One should not be overly concerned about the results of the tests for homogeneity of Level-1 variance shown in Table E.1. As previously stated, the within-class sample sizes should be at least 10 to obtain meaningful results. In addition, the more classes being evaluated the greater the chance that a significant result will be obtained even if the variance is relatively homogeneous.

The reader may notice that the degrees of freedom shown in Table E.1 are not consistent with the total number of classes, a figure that can be obtained from Table E.2. This is a result of excluding classes that were represented by only one student and for which no variance data were obtainable.

A case could be made that the natural logarithms of the within-class residual standard deviations shown in Figures E.1 through E.3 are relatively normally distributed, although several outliers do appear in the boxplots for the Applied Mathematics and Applied Technology test data. One class accounted for an outlier in both boxplots, and one individual within this class appeared to be the cause of the "outlier" status. The student took both the Applied Math and Applied Technology tests and scored below the minimum competency cutoff. The student also had an extremely high ITED score (in the 90s) coupled with a very low GPA; a very unusual situation for this data set. The data were verified through the school as correct, so incorrect data entry was not a problem here.

The other outliers were classes that had essentially inverse relationships between the predictor variables (GPAs and ITED scores) and Work Keys scores.

Figures E.4 through E.6 are included to provide the reader an idea of how the data appear; however one should remember that these plots were generated with data that did not meet the within-class sample size requirements of 10 and are therefore of little diagnostic use.

Figures E.7 through E.9 are included as a check of the appropriateness of the linear regression function and also to examine whether the variance of the error terms is constant. The four fitted intercept values for the Applied

Mathematics and Applied Technology data sets are obtained from the four possible combinations of the two Level-2 dichotomous variables associated with curricula type and relevancy. The relevancy variable did not show up as significant for the Reading for Information data set, so only two fitted intercept values were used--one for applied courses and one for traditional courses.

APPENDIX F: POWER OF TESTS

Overview

Two kinds of errors, α and β, may occur when testing hypotheses. If the null hypothesis is rejected when it is true, then a Type I (α) error has occurred. If the null hypothesis is not rejected when it is false, then a Type II (β) error has occurred. Power, or 1-β, is the probability of correctly rejecting the null hypothesis. The levels of significance, or *p-values*, given in the Student's t-test tables, Wilcoxon signed-rank tables, and HLM tables in Chapter 4 refer to a Type 1 error. Researchers may also wish to consider the power of the tests. Neter et al. (1990, pp. 74-75) discuss the procedures for obtaining the power of tests on regression coefficients using charts of the power function of the t-test. These charts are included in Appendix A of Neter et al. (1990, pp. 1138-1139).

Table F.1 contains rough estimates of the power of the tests. One additional degree of freedom is removed since the actual variance of the error terms is unknown and the sample standard error is used in the calculations.

Table F.1. Power of tests

Work Keys Test	Table	t-test statistic	df	power
Applied Mathematics	4.20	-5.333	11	99.6%
Applied Mathematics	4.31	5.960	60	99.99%
Applied Technology	4.36	5.446	34	99.93%
Reading for Information	4.38	3.354	36	88%

The Wilcoxon signed-rank method does not lend itself to simple estimation of the power of the test; the form of the power function is complex and its use is beyond the scope of this investigation. A discussion of nonparametric power functions and their properties may be found in Randles and Wolfe (1979).

REFERENCES

ACT Center for Education and Work. (1995). Making the Grade: Keys to Success on the Job in the 90's [Brochure]. Iowa City, Iowa: Author.

American College Testing Program. (1994, December 1). Work Keys: Preliminary Psychometric Handbook and Summary [Draft]. Iowa City, Iowa: Author.

American College Testing Program. (1996, May 10). Work Keys: Validity Supplement [Draft]. Iowa City, Iowa: Author.

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (1985). Standards for Educational and Psychological Testing. Washington, D.C.: American Psychological Association.

American Vocational Association. (1990). The AVA Guide to the Carl D. Perkins Vocational and Applied Technology Education Act of 1990. Alexandria, Virginia: Author

Bennett, C.A. (1926). History of Manual and Industrial Education Up to 1870. Peoria, Illinois: Chas. A. Bennett Co., Inc.

Bennett, C.A. (1937). History of Manual and Industrial Education: 1870-1917. Peoria, Illinois: The Manual Arts Press.

Boesel, D., and McFarland, L. (1994). National Assessment of Vocational Education: Final Report to Congress, Volume I, Summary and Recommendations (Department of Education Publication No. OR 94-3502-I). Washington, DC: U.S. Government Printing Office.

Bryk, A.S., and Raudenbush, S.W. (1992). Hierarchical Linear Models: Applications and Data Analysis Methods. Newbury Park, California: SAGE Publications, Inc.

Bryk, A.S., Raudenbush, S.W., and Congdon, R.T. (1996). HLM™: Hierarchical Linear and Nonlinear Modeling with the HLM/2L and HLM/3L Programs. Chicago, Illinois: Scientific Software International, Inc.

Carson, C.C., Huelskamp, R.M., and Woodall, T.D. (1993). Perspectives on Education in America: An Annotated Briefing. Journal of Educational Research. May/June, 293-294.

Crocker, L. and Algina, J. (1986). Introduction to Classical and Modern Test Theory. Orlando, Florida: Harcourt Brace Jovanovich College Publishers

Cronbach, L.J., & Snow, R.E. (1977). Aptitudes and Instructional Methods. New York: Irvington Publishers, Inc.

Cronbach, L.J., & Webb, N. (1975). Between-Class and Within-Class Effects in a Reported Aptitude x Treatment Interaction: Reanalysis of a Study by G. L. Anderson. Journal of Educational Psychology, 67 (6), 717-724.

Deming, W.E. (1986). Out of the Crisis. Cambridge, Massachucetts: Massachusetts Institute of Technology, Center for Advanced Engineering Study.

Dugger, J. and Johnson, D. (1992). A Comparison of Principles of Technology and High School Physics Student Achievement Using a Principles of Technology Achievement Test. Journal of Technology Education, 4 (1), 19-26.

Dugger, J. and Meier, R. (1994). A Comparison of Second-Year Principles of Technology and High School Physics Student Achievement Using a Principles of Technology Achievement Test. Journal of Technology Education, 5 (2), 5-14.

Goldberger, S. and Kazis, R. (1996). Revitalizing High Schools: What the School-to-Career Movement Can Contribute. Phi Delta Kappan, 77.(8), 547-554.

Gray, K. (1996). The Baccalaureate Game: Is It Right for All Teens? Phi Delta Kappan, 77.(8), 528-534.

Grieve, T.A. (1990). Evaluation of the Applied Academics Options Program for Business Students of Greene County Career Center. M.Ed. project, University of Dayton, Ohio.

Grubb, W.N. (1996). The New Vocationalism: What It Is, What It Could Be. Phi Delta Kappan, 77.(8), 535-546.

Hall, D. (1989). Principles of technology: A summative evaluation of student achievement in the first year. Doctoral dissertation, Iowa State University, Ames.

Hartoonian, M., and Van Scotter, R. (1996). School-to-Work: A Model for Learning a Living. Phi Delta Kappan. April, 555-560.

Isaac, S. & Michael, W.B. (1990). Handbook in Research and Evaluation (2nd ed.). San Diego, CA: EdITS

Iversen, G.R. (1991). Contextual Analysis. Newbury Park, California: SAGE Publications, Inc.

Johnson, R.A., and Wichern, D.W., (1992). Applied Multivariate Statistical Analysis. Englewood Cliffs, NJ: Prentice Hall.

Kandel, I.L., (1958). Philosophical Theories of American Education. In G.Z.F. Bereday & L. Volpicelli (Eds.), Public Education in America: A New Interpretation of Purpose and Practice. New York: Harper & Brothers.

Limback, E.R. & Rosa, B. (1996). Applied Academics: Relevant Education. National Business Education Yearbook (pp. 148-153). Reston, VA: National Business Education Association.

Madaus, G.F., Scriven, M., & Stufflebeam, D.L. (Eds.) (1983). Evaluation Models: Viewpoints on Educational and Human Services Evaluation. Boston: Kluwer-Nijhoff Publishing.

Mead, R. (1988). The Design of Experiments: Statistical Principles for Practical Application. Cambridge, Great Britain: Cambridge University Press.

Merton, R.K., Sills, D.L., & Stigler, S.M. (1984). The Kelvin Dictum and Social Science: An Excursion into the History of an Idea. Journal of the History of the Behavioral Sciences, 20. October, 319-331.

National Center on Education and the Economy (1990). America's Choice: High Skills or Low Wages!. Rochester, NY: Author

Neter, J., Wasserman, W., & Kutner, M.H. (1990). Applied Linear Statistical Models (3rd ed.). Boston: Irwin.

Norušis, M.J. (n.d.). SPSS® 6.1: Guide to Data Analysis. Englewood Cliffs, New Jersey: Prentice Hall

Pedhazur, E.J. (1982). Multiple Regression in Behavioral Research: Explanation and Prediction (2nd ed.). New York: CBS College Publishing.

Randles, R.H. & Wolfe, D.A. (1979). Introduction to The Theory of Nonparametric Statistics. New York: John Wiley & Sons

Resnick, L.B. & Wirt, J.G. (1996). Linking School and Work: Roles for Standards and Assessment. San Francisco: Jossey-Bass Publishers

Secretary's Commission on Achieving Necessary Skills. (1991). What Work Requires of Schools: A SCANS report for America 2000. Washington, DC: U.S. Department of Labor.

Shaw, M.C. (1994). The Administrative Steps for Implementing Applied Academics. National Business Education Yearbook (pp. 23-34). Reston, VA: National Business Education Association.

Snedecor, G.W. & Cochran, W.G. (1989). Statistical Methods (8th ed.). Ames, IA: Iowa State University Press.

Statistical Sciences (1995). S-PLUS Guide to Statistical and Mathematical Analysis, Version 3.3. Seattle: StatSci, a division of MathSoft, Inc.

Wang, C. & Owens, T. (1994). Multiple Approach to Evaluating Applied Academics. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, 1994. (ERIC Document Reproduction Service No. ED 378199).

Wang, C. & Owens, T. (1995). The Boeing Company Applied Academics Project Evaluation: Year Four. Evaluation Report. Portland, Oregon: The Northwest Regional Educational Laboratory. (ERIC Document Reproduction Service No. ED 381892).

Weisberg, H. I. (1979). Statistical adjustments and uncontrolled studies. Psychological Bulletin, 86. 1149-1164.

Wenrich, R.B., Wenrich, J.W. & Galloway, J.B. (1988). Administration of Vocational Education. Homewood, Illinois: American Technical Publishers, Inc.

ACKNOWLEDGMENTS

To all the people who, directly or indirectly, helped me reach this milestone--Thank you!

The following people and groups need special recognition:

- Other members of the project team; Dr. John Dugger, Dr. Oscar Lenning, Dr. Mandy Lively, Ms. Mari Kemis, and Ms. Andrea Wright.
- Members of the Iowa Department of Education's Bureau of Technical and Vocational Education; specifically Jerda Garey, Roger Foelske, Vic Lundy, and Laverne Hueholt.
- Members of the author's doctoral committee; Dr. John Dugger, Dr. Douglas Gemmill, Dr. Fred Lorenz, Dr. John Van Ast, and Dr. Roger Smith.
- All the guidance councilors, school administrators, regional coordinators, teachers, and students that took part in the study.
- The Work Keys staff at ACT.
- And most of all, my family, for sacrificing to help me reach a lifelong goal.